

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 15:14:14 ; Search time 1962.38 Seconds  
(without alignments)  
11547.332 Million cell updates/sec

Title: US-09-484-964-1  
Perfect score: 1465  
Sequence: 1 cgggaagattgttaaccc.....gataatggcagtcagcc 1465

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_bal: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pi1: \*  
13: gb\_pi2: \*  
14: gb\_pi3: \*  
15: gb\_pi4: \*  
16: em\_bal: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rod: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rod: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pi1: \*  
48: em\_pi2: \*  
49: em\_pi3: \*  
50: em\_pi4: \*  
51: em\_pi5: \*  
52: em\_pi6: \*  
53: em\_pi7: \*  
54: em\_pi8: \*  
55: em\_pi9: \*  
56: em\_pi10: \*  
57: em\_pi11: \*  
58: em\_pi12: \*  
59: em\_pi13: \*  
60: em\_pi14: \*  
61: em\_pi15: \*  
62: em\_pi16: \*  
63: em\_pi17: \*  
64: em\_pi18: \*  
65: em\_pi19: \*  
66: em\_pi20: \*  
67: em\_pi21: \*  
68: em\_pi22: \*  
69: em\_pi23: \*  
70: em\_pi24: \*  
71: em\_pi25: \*  
72: em\_pi26: \*  
73: em\_pi27: \*  
74: em\_pi28: \*  
75: em\_pi29: \*  
76: em\_pi30: \*  
77: em\_pi31: \*  
78: em\_pi32: \*  
79: em\_pi33: \*  
80: em\_pi34: \*  
81: em\_pi35: \*  
82: em\_pi36: \*  
83: em\_pi37: \*  
84: em\_pi38: \*  
85: em\_pi39: \*  
86: em\_pi40: \*  
87: em\_pi41: \*  
88: em\_pi42: \*  
89: em\_pi43: \*  
90: em\_pi44: \*  
91: em\_pi45: \*  
92: em\_pi46: \*  
93: em\_pi47: \*  
94: em\_pi48: \*  
95: em\_pi49: \*  
96: em\_pi50: \*  
97: em\_pi51: \*  
98: em\_pi52: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	1286.6	87.8	1514 9	AR087051 Sequence
2	1286.6	87.8	1514 9	HS083117 Human sentr
3	1120.2	76.5	1223 97	HS061397 Human ubiqn
4	1120.2	76.5	1227 91	BC006462 Homo sapi
5	1079	73.7	160939 76	AC079354 Homo sapi
6	1069.2	73.0	155913 81	AL513282 Homo sapi
7	1069.2	73.0	167440 80	AL359981 Homo sapi
8	1000.2	68.3	1017 97	HS038784 Human ubiqn





Db	1184	ATTAAATAAACATATTGTGCACAAATTCGTACTAATGCTTATTATTATTATTATGTA	1243
QY	1195	gaagaaataaaatgcctaaataagttctctgcataatactgynaattgcacatg	1254
Db	1244	GAATAAAAAATTCCTTAATAATAGGTTTCTTCGATTAATACTGGAATTCACATG	1309
QY	1255	acaattttcttcctcattactgcagaagatgtaattgaattcttggagcctgaagt	1314
Db	1304	ACAATTTTCTTCATTACTGACAGGATGATGTAATGACTTTGGAGCACTGAAAGT	1363
QY	1315	tactgaagtcgcttcgaaatcaagaagatttaataaagccacataacotttaatacca	1374
Db	1364	TACTGAAGTGCCTCTGAATCAAGAGATTAAATTAAGGCCAATACOTTTTAATACCA	1423
QY	1375	gtgttcgtgtttttttaaaacttgatatcccgatagtgcatattgatacagttac	1434
Db	1424	GTGTTCGTGTTTTAAAAACCTTGATATCCCGTATGTGCATATTGATACAGGTACC	1483
QY	1435	caatcatgttgataaattggcattgcacgc	1465
Db	1484	CAATCATGTTGGATAAATGGGCAATGCCACGC	1514

RESULT	3
LOCUS	HSU61397
DEFINITION	Human ubiquitin-homology domain protein P1C1 mRNA
ACCESSION	U61397
VERSION	U61397.1
KEYWORDS	GI:1518693
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	PRI 14-JAN-1997
FEATURES	complete cds

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 1223)  
Boddy,M.N., Howe,K., Etkin,L.D., Solomon,E. and Freemont,P.S.  
Ptc 1, a novel ubiquitin-like protein which interacts with the p

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

JOURNAL  
Oncogene 13 (5), 971-982 (1996)  
MEDLINE 9604063-3-3

REFERENCE	AUTHORS	TITLE
2 (bases 1 to 1223)	Howe, K., Boddy, M.N., Etkin, D.E., Solomon, E. and Freemont, P.S.	Direct Submission

**FEATURES**  
**JOURNAL**  
Submitted (20-JUN-1996) Somatic Cell Genetics, Imperial Can  
Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK  
Location: Joun1:51:1

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
146. .451
CDS

```

BASE COUNT	371 a	209 c	268 g	375 t
ORIGIN	/product="ubiquitin-homology domain protein PIC1" /protein_id="AA040388.1" /db_xref="GI:1518694" /translation="MSDQEAKPSTEDLGCKBGEYTKLVIGDSSSEIHFFVKMTTHI KKLKESTCQKRGVPMNSLRFLLFEGORIALNHNTPKELGMEDEEDVIEVYOEGTGHSTV"			

Query Match	76.5%;	Score 1120.2;	DB 97;	Length 1223;
Best Local Similarity	99.5%;	Pred. No. 1.2e-206;		
Matches 1155; Conservative	0;	Mismatches 3;	Indels 3;	Gaps 3

**Oy**    2   ggagaggatattgtaaaccccgagcgaagtctgtaccacgaagcgcgtgtgtgcgga    61  
| | | | |  
**Db**    60   GGAAGGATTTGTAAACCCCGAGCCGGTTCCTTACCACGAGCCCTGCTGCAGA    119  
| | | | |  
**Oy**    62   gaacccgggtgaagcaaccgcatcatgtctcacgagagcaaaccttcaactgag    120  
| | | | |

Db	120	GACCCCGGGTGAAGGCCACCGTCATCAGTCTGCACGAGGACCAAAACCTTCAACTGAGG	179
OY	122	acttgggggataaagaagaaaggatgatatatbaactcaaatgcatcttgacaagataga	181
Db	180	ACTTTGGGGGATTAAGAAGGAGGGTGATATATTTAACTCAAAAGTCATTGGACAGGATGCA	239
OY	182	gtgaagttcaactcaagaagtgaaatgaaacaacactctaaagaactcaagaatcact	241
Db	240	GTCAGATTCACTTCCAAAGTGAATGACAAACACATCTCAAAACCTCAAGAATCATACT	299
OY	242	gtcaaaagacaagggtgttccaatgaaatctcaactagggttctctcttggaaggtcaagaa	301
Db	300	GTCAAAGACAGGGGTGTCCAAATGATATCACACAGTTTCTCTTGAAGGTCAGGAATTG	359
OY	302	ctgatatatatactcccaaaagacatcggaatgaggaagaagaatgtaattgaattalc	361
Db	360	CTGATATATATCTCCAAAGAACCTGGGAATGGAGGAAGAAGATGATTAAGATTATTC	419
OY	362	aggaacaaacgggggggcatccaacagtttagatatcttttctttttctttcc	421
Db	420	AGGAACAACGGGGGGCTCATTCACAGTTTGAATTTCTTTATTTTCTTTTCC	479
OY	422	tcaatccttttattttaaanaatgacttcttctgttaatggtgtggtcaaaacggaatg	481
Db	480	TCAATCCCTTTTATTTTAAAAATAGTCTTTTGTATGGTGCTCAAAACGGAATTG	539
OY	482	aaaactggaccccactctcttgaaacactcggtaatttgaattctagtgctcaatctc	541
Db	540	AAAACCTGGCACCCCATCTCTTTGAACACTCTGGTAATTTGAATTTAGTCTCATTAATTC	599
OY	542	attattgttgtttcattcattgtctgcatlltttggtagcaaacctcagtcacctcattc	601
Db	600	ATTATTTGTTTGTTCATTTGTCATTTGTCATTTTGGTATCAAGCCCTCAGTCCCTCATATT	659
OY	602	accctctcctttttaaanaattagctgtgcaacagagagtgtaaccttttcaagacattgca	661
Db	660	ACCCTCTCCTTTTAAATTAATTAAGCTGTGCACAGAGAGTCACTTTTTCAGCAATTTGCA	719
OY	662	ttttcaagcctgtgtgtagtaataagaatcgcaacatgcaatgttcatatgaacttcca	721
Db	720	TTTTGAGCCTTGCGTGATTAATAATAGATCGCAATGCAAGTGTTCATATATGACTTTCCA	779
OY	722	attggccctgtgtgtcagatgtgatacttaactctctgagctgtgacttcaagtgag	779
Db	780	ATTGGCCCTGTAGTTTCAAGCATGTGATTACTTCACTCTCTGGACTGTGACTTTCAGTGGG	839
OY	780	agatgaaagcttttcagaagactbaactgtggaanaatgaccttctaacttgaagct	839
Db	840	AGATGGAAGTTTTCAGAGAACTGAATGTGGAAAAATGACCTTCTTAACTTGAAGCT	899
OY	840	acttttaaatlttgagggtctggaaccaaaagaaaggaataltcaagttgaatgcaagtg	899
Db	900	ACTTTTAAAAATTTTGAAGGCTGTGGCCAAAAGAGAGATATCAGGTTGAACTCAAGATG	959
OY	900	acaagataagtgagagaaatgacttaactccaagaatgagtttaacttgaagaanaagcatt	959
Db	960	ACAAGATAAGGAGAGATTAATGACTTAATCCAAAGATGGCTTCACTGAAGAAAAAGCAATT	1019
OY	960	taagattttttaaanaatctgttcaagaagaatcccaagaaagttctaatttcaatgcaat	1019
Db	1020	TAAATTTTAAAAAATCTGTGCAGAAAGATCCCAAGAAAGTTCTAATTTTCAATGCAAT	1079
OY	1020	taataaagctatacatgcaagaatgaaatacacaagaacatgctctcttttggaatttatt	1079
Db	1080	TAAATTAAGCTATACATGCACAATAATGAATCAACACAGAACCTCTCTTTTTCATTTATTT	1138
OY	1080	tgtaacttttggcctggagatatgtttaaagtgacatgtgctgtacagacttcaataa	1139
Db	1139	TGTACTTTTGGCCTGGGATATGGGTTTTAATTTGACATGTGTCTACACACTTCATTTAA	1198
OY	1140	aataacaataattgttccaana	1160
Db	1199	ATTAAACAATATTGTAAAAA	1219

Weyman, Wm. 2000 model 1762



RESULT 4  
 BC006462 1227 bp mRNA PRI 13-APR-2001  
 LOCUS Homo sapiens, ubiqutin-like 1 (sentrin), clone MGC:2093, mRNA,  
 complete cds.  
 DEFINITION BC006462  
 accession BC006462.1 GI:13623670  
 VERSION MGC.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 1227)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1350  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 Info@bcsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeay, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAL Plate: 8 Row: f Column: 12.  
 FEATURES  
 source location/Qualifiers  
 1..1227  
 /organism="Homo sapiens"  
 /db\_xref="LocusID:7341"  
 /db\_xref="taxon:9606"  
 /clone="MGC:2095 IMAGE:3546125"  
 /tissue\_type="Placenta, chorioncarcinoma"  
 /clone\_id="NIH-MGC-21"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 132..437  
 /codon\_start=1  
 /product="ubiquitin-like 1 (sentrin)"  
 /protein\_id="AAH06462.1"  
 /db\_xref="GI:13623671"  
 /translation="MSDOEAKPSTEDGDKKEGEYIKLYIGDSSIEHFVKVMTTHL  
 KTKLESYCOHQGVPNMSIRLFEQGRADNHTPEKLEMEDEYIEVYOEDTGHSTV"  
 BASE COUNT 386 a 208 c 259 g 374 t  
 ORIGIN  
 Query Match 76.5%; Score 1120.2; DB 91; Length 1227;  
 Best Local Similarity 99.5%; Pred. No. 1.2e-206;  
 Matches 1155; Conservative 0; Mismatches 3; Indels 3; Gaps 3;  
 QY 2 gggaaggttgtaaacccggagagaggttcgtctacccggagagccctgttgcgga 61  
 Db 1 |  
 QY 46 GAGAGATTGTATAACCCGAGGAGGTTCTGTTACCCGAGGCCCTGCTGTCGGA 105  
 Db 1 |  
 QY 62 gaccccggtgaagccacgctcatcgtctgacagagagcaaaccttcaactgag 121  
 Db 1 |

Db 106 GACCCCGGGGTAAGCCACCGCTCATCTGCTGACCGAGGAGCAAAACCTTCACTGAGG 165  
 QY 122 acttgaggatagaagcaagtgtaatatataactcaaatgacttgacagagataga 181  
 Db 166 ACTTGGGGGATTAAGAGAGAGGTGAATATATTAACTCAAGATCATTTGACAGATGCA 225  
 QY 182 gtgagattcactcaagtgtaaatgacacactctcaagaactcaagaatcaact 241  
 Db 226 GTGAGATTCACTTCAAAATGATAATGACACACATCTCAAGAACTCAAGATCACT 265  
 QY 242 gtcaagaagaggtgttccaatgaatcaactcaagttctcttgagggtcagagaatg 301  
 Db 286 GTCAAGACAGGGGTTCCTCAATGATTCACCTCAGTTTCTTTGAGAGGTGAGAAATG 345  
 QY 302 ctgatacatcactccaaagaactggaaatgagagaaagatgtaattgaaatc 361  
 Db 346 CTGATTAATCATCTCAAAAGACTGGGAATGAGAGAAAGATGATGAACTTATC 405  
 QY 362 aggaacaagggggtcattcaacagtttagatatcttcta-tttttctcttc 420  
 Db 406 AGGAACAAACGGGGGTCTTCAACAGTTTATGATTTCTTTTATTTTCTTTTCC 465  
 QY 421 ctcaatcctttttaaataaataagttctcttgtaatgtgtgtcaaacgaaat 480  
 Db 466 CTCATCTCTTTTATTTTAAATAGTTCTTTGTAATGAGGTCAAAAGCAAT 525  
 QY 481 gaaacatgac 540  
 Db 526 GAAACTGGCACCCCATCTCTTTGAACATCTGTAAATTTGAATTTGATCTCTATAT 585  
 QY 541 catcttgctgttcttcacatgtgcgatcttggtgacacacacacacacacacacac 600  
 Db 586 CATTAATGTTTGTTCATTTGCTGATTTTGTGATGATCAAGCCTTCCATAT 645  
 QY 601 taacctctctttaaataaataagtgtagacagagaggtcaaccttttcaggaattgc 660  
 Db 646 TACCTCTCTTTTAAATAATACGTGTCACAGAGAGTCACTTTTTCAGAGATTCG 705  
 QY 661 atttcaagctgtgtgtgtaataatgaatgaatgaatgaatgaatgaatgaatgaat 720  
 Db 706 ATTTTGAAGCTTGTGTGTAATAATGATGATGATGATGATGATGATGATGATGAT 765  
 QY 721 aattgacctgtatgc-agcatgtatctactcaactcctgtgactgtacttcagtg 779  
 Db 766 AATTGGCCTGATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 825  
 QY 780 agatgaagttttcagagaactgaactgtgaaataatgaacttcccttaactgaact 839  
 Db 826 AGATGGAAGTTTTCAGAGACGAACTGTGAAAAATACCTTTCCTTAAGTGAAGCT 885  
 QY 840 acttttaaatgtgaggtctgacccaagaagaagaatcatcaagttgaactgaagt 899  
 Db 886 ACTTTTAAATTTTGAGGGTCTGACCCAAAGAGAGATATCAGGTTCAATCAAGATG 945  
 QY 900 acagataagtgtagtaatgaactcaactcaagaatgcttcaactgaagaagaagcatt 959  
 Db 946 ACAGATAAGGTGAGATAATGACTCACTCAAGATGAGCTTCACTGAAAAAGGATTT 1005  
 QY 960 taagatttttaaatatctgtcagaagaatcccaagaagttcaatttcaactgaat 1019  
 Db 1006 TAAGATTTTTAAATAATCTGTGAGAGATCCAGAAAAGTTTATTTTTCATTAGCAAT 1065  
 QY 1020 taataaagctacatgacagaataatgaatacaagaagaacatgctcttttagatttalt 1079  
 Db 1066 TAATAAAGCTATCATGATCAAAATGATATACACAGAACTGCTCTTTT-GATTTATTT 1124  
 QY 1080 tgaacttttgccctggatagatggttttaatgagacattgtctgacagctcaatt 1139  
 Db 1125 TGTACTTTTGGCTGTGGATATGGGTTTAAATGACATTTCTGTACAGCTTCAATTA 1184  
 QY 1140 aataaacaatattgtcaaaa 1160  
 Db 1 |









CDS

67..372  
/note="Similar to ubiquitin and to yeast Smc3p (suppressor of MRP2); Method: conceptual translation supplied by author"  
/codon\_start=1  
/protein\_id="AAC50733.1"  
/db\_xref="GI:1574948"  
/translation="MSDOEAKPSTEDIGDKKEEYIKLYIGDSSPIHKVMTHTLKKLEESYCORGVNMNSLRFLFECORLADNHTPRELMEEDVIEYIQEOTGGHSTV"  
BASE COUNT 310 a 180 c 213 g 314 t  
ORIGIN

Query Match 68.3%; Score 1000.2; DB 97; Length 1017;  
Best Local Similarity 99.6%; Pred. No. 1.8e-183;  
Matches 1013; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 22 ggagcgaagctctgcttaaccgagcgcgcgctgcygagagaccgaggtgaagccacc 81  
Db 1 ggagcgaagctctgcttaaccgagcgcgcgctgcygagagaccgaggtgaagccacc 60  
Qy 82 gtcacatgctgacccagaggaacacacacacacacacacacacacacacacacacac 141  
Db 61 gtcacatgctgacccagaggaacacacacacacacacacacacacacacacacacac 120  
Qy 142 ggctgaatataatgaactgaacacacacacacacacacacacacacacacacacac 201  
Db 121 ggctgaatataatgaactgaacacacacacacacacacacacacacacacacacac 180  
Qy 202 aaatgaac 261  
Db 181 aaatgaac 240  
Qy 262 atgaattcactcaagcttctctcttgaaggtgaagaaatgctgaatacactcccaaa 321  
Db 241 atgaattcactcaagcttctctcttgaaggtgaagaaatgctgaatacactcccaaa 300  
Qy 322 gaactgggaatggaggaagaatgagattgaattatcaagaacacacacacacacac 381  
Db 301 gaactgggaatggaggaagaatgagattgaattatcaagaacacacacacacacac 360  
Qy 382 tcaacagcttagaatctcttcttcttcttcttcttcttcttcttcttcttcttctt 441  
Db 361 tcaacagcttagaatctcttcttcttcttcttcttcttcttcttcttcttcttctt 420  
Qy 442 aaaaagctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 501  
Db 421 aaaaagctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 480  
Qy 502 ttgaac 561  
Db 481 ttgaac 540  
Qy 562 tgcgtgatttgggagatgagctgagctgagctgagctgagctgagctgagctgagct 621  
Db 541 tgcgtgatttgggagatgagctgagctgagctgagctgagctgagctgagctgagct 600  
Qy 622 taagtgtgacacagaggtgacacacacacacacacacacacacacacacacacac 681  
Db 601 taagtgtgacacagaggtgacacacacacacacacacacacacacacacacacac 660  
Qy 682 aataagaatcacaac 740  
Db 661 aataagaatcacaac 720  
Qy 741 atgtgattactcactcctgagctgagctgagctgagctgagctgagctgagctgagct 800  
Db 721 atgtgattactcactcctgagctgagctgagctgagctgagctgagctgagctgagct 780  
Qy 801 ctgaactgtggaaaatgac 860  
Db 781 ctgaactgtggaaaatgac 840

Qy 861 ggaccacaaagaagagatatacaggttgaagtcacagacacacacacacacacacacac 920  
Db 841 ggaccacaaagaagagatatacaggttgaagtcacagacacacacacacacacacacac 900  
Qy 921 actaac 980  
Db 901 actaac 960  
Qy 981 tcagaagatccac 1037  
Db 961 tcagaagatccac 1017

RESULT 9  
AC005220 163196 bp DNA PRI 01-JUL-1998  
LOCUS Homo sapiens chromosome 20, BAC clone 99 (IBNL H80), complete  
DEFINITION sequence.  
ACCESSION AC005220  
VERSION AC005220.1 GI:3282159  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,  
1 (bases 1 to 163196)  
Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.  
Sequencing of human chromosome 20  
Unpublished  
2 (bases 1 to 163196)  
Ricke,D.O.  
TITLE Large Scale Sequence Analysis and Annotation with the Sequence  
Comparison Analysis (SCAN) System  
JOURNAL Unpublished  
3 (bases 1 to 163196)  
AUTHORS Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,  
Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.  
DIRECT SUBMISSION  
Submitted (01-JUL-1998) Human Genome Center, DOE Joint Genome  
Institute, Lawrence Berkeley National Laboratory, MS 74-157,  
Berkeley, CA 94720, U.S.A.  
COMMENT Sequence submitted by:  
DOE Joint Genome Institute.  
FEATURES  
Source  
Location/Qualifiers  
1..163196  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="20"  
/clone="BAC 99"  
/chromosome="20"  
/note="IBNL H80"  
4588..4709  
/rpt\_family="MIR"  
6685..6813  
/rpt\_family="Tiggr1"  
complement(6838..9257)  
/rpt\_family="L1"  
7175..7314  
/rpt\_family="L1"  
7322..7441  
/rpt\_family="L1"  
complement(8695..9257)  
/rpt\_family="MIR2B2"  
10121..10350  
/rpt\_family="Alu"  
11057..11357  
/rpt\_family="Alu"  
13431..13702  
/rpt\_family="Alu"  
15911..15962  
repeat\_region

```

misc_feature      /rpt_family="MIR"
                  16029..16484
                  /note="99% identity B54864 BAC end sequence"
repeat_region     /rpt_family="Alu"
                  complement(16891..17025)
repeat_region     /rpt_family="Alu"
                  complement(17221..17682)
repeat_region     /rpt_family="L1"
                  complement(18500..18756)
misc_feature      /rpt_family="L1"
                  complement(18909..19161)
repeat_region     /note="GRAIL 2 excellent exon, frame 1"
                  19640..19943
                  /rpt_family="Alu"
                  complement(22025..22314)
repeat_region     /rpt_family="Alu"
                  complement(22903..23020)
misc_feature      /note="GRAIL 2 excellent exon, frame 1"
                  complement(23247..23567)
repeat_region     /rpt_family="Alu"
                  complement(24441..24514)
misc_feature      /note="GRAIL 2 excellent exon, frame 0"
                  25506..25525
                  /note="(T)20"
                  /rpt_type=tandem
repeat_region     /rpt_unit="r"
                  complement(25538..25679)
repeat_region     /rpt_family="MIR"
                  complement(26920..27577)
repeat_region     /rpt_family="LTR8"
                  complement(28412..28563)
repeat_region     /rpt_family="MER20"
                  28734..29019
repeat_region     /rpt_family="Alu"
                  30266..30455
repeat_region     /rpt_family="THE1"
                  30789..30888
misc_feature      /note="GRAIL 2 excellent exon, frame 2"
                  31841..31972
misc_feature      /note="GRAIL 2 excellent exon, frame 0"
                  32270..32290
                  /note="(T)21"
                  /rpt_type=tandem
repeat_region     /rpt_unit="r"
                  32702..32860
misc_feature      /note="GRAIL 2 excellent exon, frame 1"
                  33268..33474
misc_feature      /note="GRAIL 2 excellent exon, frame 0"
                  33836..33863
                  /note="(GT)14"
                  /rpt_type=tandem
repeat_region     /rpt_unit="Gr"
                  35158..35265
misc_feature      /note="GRAIL 2 excellent exon, frame 0"
                  complement(41855..42137)
repeat_region     /rpt_family="Alu"
                  43253..43465
misc_feature      /note="GRAIL 2 excellent exon, frame 1"
                  complement(50487..50600)
repeat_region     /rpt_family="L1"
                  50605..50855
repeat_region     /rpt_family="L1"
                  51461..51822
repeat_region     /rpt_family="THE1"
                  complement(54343..54856)
repeat_region     /rpt_family="LTR12"
                  complement(55490..55796)
repeat_region     /rpt_family="Alu"
                  55900..56027
repeat_region     /rpt_family="MIR"
                  complement(61007..61096)
misc_feature      /note="GRAIL 2 excellent exon, frame 1"
                  join(61724..61846,71254..71319,75055..75261)
prim_transcript   /standard_name="af74a01"

```

```

repeat_region     /note="100% identity dbEST:AA644590"
                  complement(62247..62746)
repeat_region     /rpt_family="L1"
                  complement(65258..65541)
repeat_region     /rpt_family="Alu"
                  65587..65882
repeat_region     /rpt_family="Alu"
                  68256..68282
                  /note="(T)27"
                  /rpt_type=tandem
repeat_region     /rpt_unit="r"
                  69077..69495
repeat_region     /rpt_family="Alu"
                  69537..69618
repeat_region     /rpt_family="Alu"
                  complement(70542..70721)
repeat_region     /rpt_family="MERS"
                  72072..72255
repeat_region     /rpt_family="MIR"
                  complement(72260..72340)
repeat_region     /rpt_family="MLT1"
                  7417..74456
                  /note="(CT)20"
                  /rpt_type=tandem
repeat_region     /rpt_unit="Cr"
                  75057..75320
misc_feature      /note="GRAIL 2 excellent exon, frame 2"
                  join(75199..75323,84601..84746)
prim_transcript   /standard_name="zm96b07"
                  /note="96% identity dbEST:AA079436"
repeat_region     complement(77059..77202)
                  /rpt_family="MERS"
repeat_region     complement(77269..77551)
                  /rpt_family="Alu"
prim_transcript   join(78685..79043,88203..88578)
                  /standard_name="ym20b03"
                  /note="97% identity dbEST:HI5040"
                  79526..79590
prim_transcript   /standard_name="nc17d10"
                  /note="81% identity dbEST:AA259187"
repeat_region     complement(81547..81627)
                  /rpt_family="MIR"
repeat_region     82700..82865
                  /rpt_family="Alu"
repeat_region     83191..83347
                  /rpt_family="L1"
repeat_region     complement(83468..83594)
                  /rpt_family="MIR"
repeat_region     84144..84436
                  /rpt_family="Alu"
                  complement(84402..84774)
prim_transcript   /standard_name="n14b12"
                  /note="98% identity dbEST:AA554549"
                  complement(85065..85162)
prim_transcript   /standard_name="yw21h10"
                  /note="90% identity dbEST:H88501"
                  join(85165..85589,86249..86579,87354..87476)
prim_transcript   /standard_name="y909d02"
                  /note="97% identity dbEST:R43020"
                  complement(86451..86577)
misc_feature      /note="GRAIL 2 excellent exon, frame 2"
                  complement(87953..88445)
prim_transcript   /standard_name="zu90c12"
                  /note="99% identity dbEST:AA626177"
                  complement(89695..89866)
misc_feature      /note="GRAIL 2 excellent exon, frame 1"
                  89959..90146
repeat_region     /rpt_family="MERS"
                  92028..92411

```

Query Match 55.5%; Score 813; DB 85; Length 163196;  
 Best Local Similarity 88.0%; Pred. No. 2.8e-147;  
 Matches 1032; Conservative 0; Mismatches 105; Indels 36; Gaps 12;

QY 2 gggaaagatttgtaaaccccgagcaggtctctta-cccagagccgcgtcgtcgcg 60  
 Db 153108 GAGATGGCATTGTAAACCCAGAGGAGTCTGCTACCCGAGACGCTGCTGTCGG 153167  
 QY 61 agaccccggttgaagccacgtcatcgtctgaccagagagcaaaccttcaactgag 120  
 Db 153168 AGACTGCGAGTGAATGCCCATCCATGCTGACCTGGAGGCAAAACCTTCAACTGAG 153227  
 QY 121 gacttggggatagaagaacagtgatataataaactcaagtcattgagcagagatagc 180  
 Db 153228 CATTGGGGGATGAAGATAAAGATATTAACCTCAGGGGTTATTGGACAGATAGC 153287  
 QY 181 agtgaattccttcaaaagtgaatatacaacacatctcaagaactcaagaatctac 240  
 Db 153288 AGTGAGATTCTTTCAAAAGTAAATGACACACCTTCAAGAACTCAAGAAATCGTAT 153347  
 QY 241 tctcaagaacaggggttccaatgaattcactcaggttctcttggagggtcagagatt 300  
 Db 153348 TGTGAGAGACAGGGGCTTCCAGTGAATTCCTCAGTTCTCTTGAAGGTCAGAGAAAT 153407  
 QY 301 gctgaatacactcaaaaagacttgggaatggaggaagaagatgtgattgaattat 360  
 Db 153408 GCTGATATATCTACTCCAGAGAACTGGGAATGGAGAAAGATGATGAGTTGAGGTTAT 153467  
 QY 361 caagaaacacaggggttcaatcaacagttatagatatcttcttatttcttcttcc 420  
 Db 153468 CAGAGCAAAATCGAGGTCATTCACAGTTTACACATCTTTT-TTTTTCTCTTTTC 153525  
 QY 421 ctcaatccttttattttaaataagctcttctgtaatgtgtgtcacaacaggaatt 480  
 Db 153526 CTCAATCTTTTATTTTATTTTAAATGATCTTTTGTAGCATGATTTAAACGGAAT 153585  
 QY 481 gaaacatgacccacatctcttgaacacatctgttaattgaattctagtcattatt 540  
 Db 153586 GAAACAGCAGCATCCATCTCTTAAACATCTGTAATTTGAATTTAGTTCATTATT 153645  
 QY 541 cattattgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 600  
 Db 153646 CATTATGTTGTTTCTTCAATGATGCTGATTTCTGATCAGGCTCAGGCTCTTCATAT 153705  
 QY 601 taccctctctt-aaataatgctgcacagagaggttcaacttcttcttcttcttct 659  
 Db 153706 TGCCCTCTCTTTTAAATAAATTAATGTGACACAGAGGCACTTTTTCAGACTGTA 153765  
 QY 660 catttcaaggtctgtgtataataagatcagcaatgcaagtgtlcaatactcttc 719  
 Db 153766 CATTTCAG-----ATGATTAAGATGAGCAATGCAAGTATTCATATGACTTTC 153814  
 QY 720 caattggccctgatgtc-aggatgttacttacttcttcttcttcttcttcttcttct 778  
 Db 153815 CAGTTGGCCCTGATTTTCAACCATGATGATTTCTGATCCTCGACTATGACTTTGGTGG 153874  
 QY 779 ggaatggaagtttttcaagaacttgaactgttgaataaataagcttcttcttcttct 835  
 Db 153875 GAGATGGAAGTTTATGAGACTGAGCTGTGGAATAATGACCTCTCTTAACAACCTTGA 153934  
 QY 836 agctactttaaatttgaaggtctgagcaagaagaagaatatacaggttgaagtaac 895  
 Db 153935 ACCTATTTTAAAGATTTGAGGCTGTGACCAAAAGAAAGAGCATGAGGTT-AAAGTCAA 153993  
 QY 896 gatgacagaataag-tgaagatgaactaactcaagaatgagcttcaacttgaataaagg 954  
 Db 153994 GGTGACAGATAAGTTGAGAGTAAGACTAATCCAAAGATGGCTTCCCTTAAGAAAGA 154053  
 QY 955 catttaaga-----ttttttaaactctgtcagaagaatcccaagaagaagtcttaatttc 1040  
 Db 154054 CATTGGAAGATTTTCTTAAATCTTGTGAGAGATCCCAAAAAGTTCTGATTTTC 154113  
 QY 1011 attgacaataaagaatgatacagcaagaatgaataaacaagaacagcttcttcttla 1070  
 Db 154114 ATT-----ATCAATTAAGATTAATGCAAGAAATGAATACACAGAAATACGCTCTT---- 154165

QY 1071 gatttattgttacttcttggcctgggatat-gggtttaaatgacattgtctlacca 1129  
 Db 154166 -TTTATTGTACTTTTGGCTGGATATCGGCTTTTAAATGACATTTGCTACACCA 154223  
 QY 1130 gcttactttaaataacataattgtcaaat 1162  
 Db 154224 GCTTCATTTAAATAATATTTGTAATAATT 154256  
 RESULT 10  
 HSU67122 816 bp mRNA PRI 15-FEB-1997  
 LOCUS  
 DEFINITION Human ubiquitin-related protein SUMO-1 mRNA, complete cds.  
 ACCESSION U67122  
 VERSION U67122.1 GI:1762972  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Mahajan,R., Delphin,C., Guan,T., Gerace,L. and Melchior,F.  
 TITLE A small ubiquitin-related polypeptide involved in targeting RanGAP1  
 to nuclear pore complex protein RanBP2  
 JOURNAL Cell 88 (1), 97-107 (1997)  
 MEDLINE 97148692  
 REFERENCE  
 AUTHORS Mahajan,R.K., Guan,T., Delphin,C., Gerace,L. and Melchior,F.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-AUG-1996) Department of Cell Biology, The Scripps  
 Research Institute, 10666 N. Torrey Pines Road, La Jolla, CA 92037,  
 USA  
 FEATURES  
 source  
 CDS  
 81..386  
 /note="Conjugated post-translationally to RanGAP1;  
 ubiquitin-related protein; similar to UBL1 encoded by  
 Genbank Accession Number U61397 and GMP1 encoded by Genbank  
 Accession Number U72722"  
 /codon\_start=1  
 /product="SUMO-1"  
 /protein\_id="AAC50996.1"  
 /db\_xref="GI:1762973"  
 /translation="MSDQAKPSTEDLDGKKEGEVILKVIQDSSEIHFKVMKTHL  
 KKLKESYCOGROGVPMNSLRFLEGGRIADNHPKELGMEEDVIEVYQDTGSHSTV"  
 BASE COUNT 235 a 153 c 173 g 255 t  
 ORIGIN  
 Query Match 52.6%; Score 770; DB 97; Length 816;  
 Best Local Similarity 98.9%; Pred. No. 4.8e-139;  
 Matches 807; Conservative 0; Mismatches 5; Indels 4; Gaps 3;  
 QY 8 gatttgaaccccgagcaggttctgcttaccgaagccgctgtctggtggagagccc 67  
 Db 1 GATTTGAACCCCGAGCGAGGTTCTGCTTACCGAGCGCTGCTGCGAGACCC 60  
 QY 68 cgggtgaagccacgtcatcatgtctgacagagagcaaaccttcaactgagagattg 127  
 Db 61 CGGTAAGCCACGTCATCATCTGTGACAGAGCAAAACCTTCAACTGAGACTTGG 120  
 QY 128 gggataagaagcaggttgatatataactcaaaatgcatcttgacagagtagcagtaga 187  
 Db 121 GGGATGAAGAGGAGGTGATATTAATACTCAAAAGTCTTGGACAGGATACACTGAGA 180  
 QY 188 ttacttcaaatgtaaaatgacaacacatctcaagaactcaagaatcaatcatctgcaaa 247  
 Db 181 TTCACTTCAAAAGTGAATAATGACACACATCTCAAGAACTCAAGAATCATCTGTCAAA 240



QY 248 gacaggggtgttccaatgaatcaccacaggttctcttggagggcagagaatgtgata 307  
 |||||||  
 Db 241 GACAGGGGTGTCGAATGAATCACTCAGGTTTCTTTGAGGGTCAGAGAAATGCTGATA 300  
 QY 308 atcaactcctcaaaagacatgggaatgagagaagaatgtgatgaaatcattcaagaac 367  
 |||||||  
 Db 301 ATCATACTCCAAAGAACTGGGAATGAGAAAGATGTGATTGAAGTTATCAGGAAC 360  
 QY 368 aaagggggggcattcaacagattagattcttcttatttttttcttcttccctcaac 427  
 |||||||  
 Db 361 AAAGGGGGGGCATTCAACAGATTAGATATTCTTTTATTTTCTTTCTTCCCTCAATC 420  
 QY 428 ctttttattttttaaataagttcttcttgaatgtgtgttcaaaacggaatgtgaaaac 487  
 |||||||  
 Db 421 CTTTATTATTTTAAATAATAGTCTTTTGTATGNGTGTCAAAACGGAATGAAAAC 480  
 QY 488 ggcaccacatctcttgaacaacatctgttaattgaaatctgagtcataattatatt 547  
 |||||||  
 Db 481 GGCACCCCATCTCTTTGAAACATCTGTAAATTTGAATTTCTGTCTCATATTCATTAAT 540  
 QY 548 gttgttttcatgtgtcattttttgtgtatcaagcctcagctccctcatattacccctc 607  
 |||||||  
 Db 541 GTTGTATTTTCAATGTGCTGATTTTGGTGATCAACCCCTCATATTATTAACCTC 600  
 QY 608 tccctttaaataatcaatgtgtgacagagaagtcaccccttctcaggaattgatttca 667  
 |||||||  
 Db 601 TCCCTTTTAAATAATTTACGTGTGACAGAGAGTCCCTTTTCAGAGATTCATTTTCA 660  
 QY 668 gacttggtgtgttaataatgaatgcacacatgaatgtgtcattgaattcccaattgc 727  
 |||||||  
 Db 661 GCGCTGTGTGTAAATTAATGATGACCAATGCAATGTTCAATTAATGACTTTTCAATTGGC 720  
 QY 728 cctgatgttc-agcatgtgattacttactc-cctgactgtgacttcaagtgagagatg 785  
 |||||||  
 Db 721 CCGTATGTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780  
 QY 786 aagcttttc--agagaactgaaactgtgaaaaatga 819  
 |||||||  
 Db 781 AATTTTTCAGAAAAAAGTGAACGTGTGAAAAATGA 816  
 RESULT 11  
 AF033353 1187 bp mRNA ROD 06-FEB-1998  
 LOCUS AF033353 ubiquitin-homology domain protein (Ub1) mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF033353  
 VERSION AF033353.1 GI:2645736  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 1187)  
 AUTHORS Howe, K., Williamson, J., Boddie, N., Sheer, D., Freemont, P. and  
 Solomon, E.  
 TITLE The ubiquitin-homology gene P101: characterization of mouse (P101)  
 and human (UB1) genes and pseudogenes  
 JOURNAL Genomics 47 (1), 92-100 (1998)  
 MEDLINE 98126440  
 REFERENCE 2 (bases 1 to 1187)  
 AUTHORS Howe, K., Freemont, P.S. and Solomon, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-NOV-1997) Medical & Molecular Genetics, Guy's  
 Hospital, 8th floor, Guy's Tower, London SE19RT, UK  
 FEATURES  
 source  
 1. 1187  
 /organism="Mus musculus"  
 /strain="ICR"  
 /DB\_xref="taon:10090"  
 /chromosome="1"  
 /map="C2-3"  
 1. 1187  
 gene

CDS  
 /gene="ub1"  
 /note="P101"  
 119..424  
 /gene="ub1"  
 /codon\_start=1  
 /product="ubiquitin-homology domain protein"  
 /protein\_id="AAC39959.1"  
 /db\_xref="GI:2645737"  
 /translation="MSDPAKSTEDLDKKEGEYIKIKVIGDSSSEIHFKVKTTHL  
 KRLESYCORQGVPMNLSRFLEGEGRIDNHTPKELGMEDEYIEVYEQGGHSIV"  
 BASE COUNT 350 a 216 c 267 g 354 t  
 ORIGIN  
 Query Match 52.3%; Score 765.8; DB 94; Length 1187;  
 Best Local Similarity 84.4%; Pred. No. 3.2e-138;  
 Matches 988; Conservative 0; Mismatches 152; Indels 31; Gaps 10;  
 QY 2 gggaaagattgttaaaccccgagagagtcgtcttaaccgagcgctgtgtcgga 61  
 |||||||  
 Db 34 GGGACGGATTGTAAACCTTCAGATGAGTTCTGCTCCGGAGGCGCAGTGTGCGGA 93  
 QY 62 gaccccggttgaagccacgcatcagtcgtcagcagaagagcaaaccttcaactgag 121  
 |||||||  
 Db 94 GACCCCGGGTGAATCCA-CGTACCATGTGTGACAGGAGCAAAACCTTAACAGAG 152  
 QY 122 acttgggggataaagcaaggtgaaatataataactcaaatgcatgtgacagataga 181  
 |||||||  
 Db 153 ACTTAGCGATTAAGAAAGAGAGGAAATATTAATTAATTAATTAATTAATTAATTAAT 212  
 QY 182 gtgagattcaactcaaaagtgaataatgacaacacatctcaagaactcaagaatcatt 241  
 |||||||  
 Db 213 GTGAGATATCATTTTAAAGTGAATGACACACATCTCAAGAAACCTCAAGATCATAT 272  
 QY 242 gtcaagaagaggtgttccaaatgaatcactcaaggttctcttggaggtcagagaatt 301  
 |||||||  
 Db 273 GTCAAGACAGAGGAGTTCATTAATGATTCACACAGTTCCTTTGAAGGTGAGGAAT 332  
 QY 302 ctgtaatacatcaccacaaagaactgtggaatgagagaagaatgtgattgaagttatc 361  
 |||||||  
 Db 333 CTGATTAATCATCTCCGAAAGAACTGGGAATGAGGAAGAAAGATGATGAAGTTATC 392  
 QY 362 aggaacaaacgggggtgcatcacaacagtttagata----tcttttattttttctt 416  
 |||||||  
 Db 393 AGGAACAAACGGGGGTGCTACGACGGTTAGATTAATCTTTTATTTTATTTTCC 452  
 QY 417 ttcctcaatccttttattttaaataatgtcttcttgaatgtgtgttcaaaagc 475  
 |||||||  
 Db 453 TTCCCTCAATCTCTTTTATTTTAAATAATGTTCTTTGTAATGTGTGTTCAAAATGA 512  
 QY 476 gaattgaaactgaccccatctcttgaacatctgtaattgaaattgaaattgactca 535  
 |||||||  
 Db 513 AATTTGAATCTGCGACATCTCTTTGAA-----CATATGAATCTGATGTTGTTG 563  
 QY 536 ttatcattatgttgttcttcaatgtgtcgtgattttgtgtgataaagcctcaagccct 595  
 |||||||  
 Db 564 AATTCATTAATGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 623  
 QY 596 catattaccccttcttctttaa---aaattcgtgtgacagagaggtcacttttca 652  
 |||||||  
 Db 624 AATTAAGTCCCTTTTCCCTTTTAAAGATTTTCAATGTGTGACAGAGAGGCCACTTTTCAG 683  
 QY 653 gacattgacattcagagctgtgtgataaataagatcgacaatgcaagtgttcaataat 712  
 |||||||  
 Db 684 GACTGTGCAATTTTCAAGTTTGTGATGATAAA-AAGATGACCAATGAGGATTTCCATG 742  
 QY 713 gacttcaaatgtgcccgtgattgaatgtgattgaattgaattgaattgaattgaatt 772  
 |||||||  
 Db 743 ACCTTCAATGTGCTGTGAAGTTCCAGCATGTGTTGCTTCACTCTGACATGATTTT 802  
 QY 773 cagtgagagatgagatttttccagagaactgaactgtgaaataatgactttccttaact 832  
 |||||||  
 Db 803 CAGTGGAGATGGAATTTTTCAGAGAACTGAACGTGTG-AAATGACCTTTCCACACT 861

QY 833 tgaagctactttaattgagggtcgtgacacaaagaagaatcaggttaagt 892  
 Db 862 TGAAGCTACTTTTAAATCTCGGGCTGACCAAAAGAACATCAT -GTTTGTAGT 919  
 QY 893 caagatgacagataagtgagagatgaatgaactccaagaatgctcactgaagaa 952  
 Db 920 CAAGATGACAGATACAGAGACAGACAGTAACTCCAAAGTGGCTTCATCGAGAGAA 979  
 QY 953 ggcatttaagatttttaaaaatctgtcagaagaatccagaagaatttcaat 1012  
 Db 980 AGAAGTCTCTTGACAGACAGATATGTCAGAGATCCAGAAAGTTCTAATGTTCA 1039  
 QY 1013 taaga--attaataagctatacatcagaatgaatacacagaacactgcctttt 1069  
 Db 1040 CAGAGTTATTAATTAAGTTACTATACAGAGATGTCAGACACAGA-----CACTGCTTT 1094  
 QY 1070 agatttatttctacttttgctgagatagggtttaatgacatgtctgtacca 1129  
 Db 1095 TGATTCTGTTTGTACTTTTGGCTGGACATGGTTTCAAGGACATCGTGTGTACA 1154  
 QY 1130 gcttcataaataacaatatgttcaaaa 1160  
 Db 1155 GCTTCATTAAATAACAATATTGTAAAA 1185

RESULT 12  
 AC084066/c  
 LOCUS  
 DEFINITION Mus musculus clone RP23-321D1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 29  
 unordered pieces.  
 ACCESSION AC084066.1 GI:10799415  
 VERSION AC084066.1  
 KEYWORDS HTG; HTGS; PHASE1.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 235411)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Mouse  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 235411)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----  
 Project Information  
 Center Project Name: 2351294  
 Center clone name: RPCI-23\_321D1  
 -----

Summary Statistics  
 Consensus quality: 214207 bases at least Q40  
 Consensus quality: 223053 bases at least Q30  
 Consensus quality: 225208 bases at least Q20  
 Estimated insert size: 200000; pulse field gel estimation  
 Estimated insert size: 232611; sum-of-contigs estimation  
 Quality coverage: 11.56 in Q20 bases; pulse field gel estimation  
 Quality coverage: 9.94 in Q20 bases; sum-of-contigs estimation  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 29 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1210: contig of 1210 bp in length

FEATURES  
 source  
 1.235411  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-321D1"  
 /clone\_lib="RPCI mouse BAC library 23"  
 BASE COUNT 62132 a 54445 c 54225 g 61806 t 2803 others  
 ORIGIN  
 1211 1310: gap of unknown length  
 1311 3240: contig of 1930 bp in length  
 3241 3340: gap of unknown length  
 3341 4504: contig of 1164 bp in length  
 4505 4605: gap of unknown length  
 4606 6172: contig of 1568 bp in length  
 6173 6273: gap of unknown length  
 6273 7354: contig of 1082 bp in length  
 7355 7455: gap of unknown length  
 7455 8626: contig of 1171 bp in length  
 8626 8726: gap of unknown length  
 8726 10114: contig of 1389 bp in length  
 10115 10214: gap of unknown length  
 10215 12091: contig of 1877 bp in length  
 12092 12191: gap of unknown length  
 12192 14113: contig of 1922 bp in length  
 14114 14213: gap of unknown length  
 14213 15287: contig of 1073 bp in length  
 15287 15386: gap of unknown length  
 15386 17839: contig of 2453 bp in length  
 17839 17939: gap of unknown length  
 17939 20266: contig of 2327 bp in length  
 20267 20366: gap of unknown length  
 20367 22001: contig of 1635 bp in length  
 22002 24307: gap of unknown length  
 24307 24407: contig of 2206 bp in length  
 24407 26405: gap of unknown length  
 26405 26505: contig of 1998 bp in length  
 26506 29691: contig of 3186 bp in length  
 29692 29791: gap of unknown length  
 29791 34705: contig of 4914 bp in length  
 34705 34805: gap of unknown length  
 34805 39749: contig of 4944 bp in length  
 39749 39849: gap of unknown length  
 39849 45296: contig of 5447 bp in length  
 45297 45397: gap of unknown length  
 45397 51477: contig of 6080 bp in length  
 51477 51577: gap of unknown length  
 51577 59009: contig of 7432 bp in length  
 59009 59109: gap of unknown length  
 59109 66218: contig of 7110 bp in length  
 66218 66319: gap of unknown length  
 66319 76779: contig of 10460 bp in length  
 76779 76879: gap of unknown length  
 76879 80898: contig of 21220 bp in length  
 80898 98198: gap of unknown length  
 98198 113987: contig of 15789 bp in length  
 113987 114087: gap of unknown length  
 114087 132115: contig of 18028 bp in length  
 132115 132215: gap of unknown length  
 132216 160184: contig of 27969 bp in length  
 160185 160284: gap of unknown length  
 160285 186798: contig of 26414 bp in length  
 186799 235411: gap of unknown length  
 235411 48613: contig of 48613 bp in length.

Query Match 50.8%; Score 743.6; DB 76; Length 235411;  
 Best Local Similarity 83.6%; Pred. No. 7e-134;  
 Matches 971; Conservative 0; Mismatches 159; Indels 32; Gaps 10;  
 QY 12 tgaataaccgagagaggtctgcttaaccgagcgctgctgtgagagaccccgag 71  
 Db 113180 TGAATAACTCAGAGTGTCTGCTGCGGAGCGCAGTGTGTGACAGAGACCCCGGG 113121

Qy	72	tgaagcaacgfcatactatgctcgaccagagggcaaaaccttcaacttgaggacttggggg	131
Db	113120	TGAATCCAGTCGC--CATCTCTGCACGAGGAGCAAAACCTTCAACTGACGACTT	113063
Qy	132	taagaagcaagtgatatacatlaaactcaaaagtcatttgacaagatagcagtgagatca	191
Db	113062	TAAAGAAAGAAAGGAGATACATTAAACTCAAAAGTTATGTGACAGCAGACGATGAGATACA	113003
Qy	192	cttcaaaagtgaanaatgacacacatctcaagaaactcaagaatcatatcgttcaagaaca	251
Db	113002	TTTTCAAGTGAATAATGACAAACACATCTCCAAAGAACTCAAAAGATCATCTCAAAAGACA	112943
Qy	252	gggtgttccaatgaatcaactcaaggtctctcttgaaggtcagaagaattgtctgataca	311
Db	112942	GGGGATTTCATGATTAATTCACCTACGCTTCTCTTTGAAGTGCAGAAATATGCGTGAATACA	112883
Qy	312	tactccaaagaactgggaatggagaagaagtatgattgaagtattcaagaaacaac	371
Db	112882	TACTCCAAAGAAGAACTGGGAGATGGAGAAAGAAAGATGTGATGAAGTTTATCGAGAAACAAC	112823
Qy	372	gggggggcatccaagagtttagata-----tctctttaaatttttcttcttccctcaat	426
Db	112822	GGGGGGGTACCTCGAGGGTTAGTAATTCCTTTTATTTTATTTTCTTTCCTCCATAT	112763
Qy	427	ccttttattttttaaanaatagcttcttcttgtaaagtgtgttcaaaagc--gaattgaana	485
Db	112762	CGTTTTTATTTTAAAAATAGTTCTTTTGTAATGTGGTGTTCAAAAGATGAATTAATGAATA	112703
Qy	486	ctggacaccccatctcttgganaaatctcgtgaaatttgaattcagtcagtcattatcatca	545
Db	112702	CTGGACACCCCATCTCTTGAAA-----CATATGAATTTAGTGTTCATATTAATTCATTA	112652
Qy	546	tgtgttcttctcatgtgctgagatlttltgtgatacaagcctcagtcctccatlatcaccc	605
Db	112651	TGTGTTGTTTTTGTGTGTGTCGATATTTTGGTATGACAGCCACACCTCCCTTAATACGTGCC	112592
Qy	606	tctcctttttaa--aaattaagtgtgacagagaggttcaccttcttcaagagcattgcat	662
Db	112591	TTTTTTCCTTTTAAAGAAATTTCAATGTGTGCACAGAGAGGCCACCCCTTTTATGAGCATGTGCAT	112532
Qy	663	tttctagccttctgtgtgatataaataagatcgcagcaatgcgaatgtgtcatalatgactccaa	722
Db	112531	TTTTCAAGTTTGTGATGATTAATA--AAGATCGACCAATGGGACCTTCCATAGACCTTTCAAT	112473
Qy	723	ttagccctgatagttcaagcatgatacttaactcctctgacgtgtgacttcaagtggagaga	782
Db	112472	TGGCTCTGAAGTTCACAGCATGTGGTGGCTTCACTCTCGGACGTGATTTTCACTGGGAGA	112413
Qy	783	tgggaagtttttcaagagaactgaactgttgaaaaaataatgaccttcttcaacttgaagctact	842
Db	112412	TGGGAATTTTTTACAGAACTGAACTGAAGTGG--AAAAATGACTTTTCTCTCAAGCTTGAAGTACT	112354
Qy	843	ttttaaatttggaggtcttggagccaaaagaaggaataatacagtttgaagtcaagaatgaca	902
Db	112353	TTTTAAATCTGGGGGCTGAGACCAAAAGAAAGAAATCATCAT--GTATTGTATGCAAGATGACA	112296
Qy	903	gataaggtgaagatatagtcaactcaactcgaagaatgttcttaacttgaagaagaagcatltaa	962
Db	112295	GATATCAGTATGAGCAACACACTTAATCTCCAAAGGTGGCTTCACTGAGAGAAAGAAAGATGTC	112236
Qy	963	gattttttaaanaatctgtcagagaagaatcccgaaagaagtttcaatttccattagaca--at	1019
Db	112235	TTTGAGCAAGACAGTATTTGTTCAGAGATCCAGCAAAAGTTTAAATGTTTCACTCAGCAGTTAT	112176
Qy	1020	taataaagctatacatatgcagaataatgaatacaacgaagaacatgcctcttctttagatttatt	1079
Db	112175	TAAATAAAGTTTACATATGAGAAAGTATGACCAACAACA-----CACTGCTTTTGAATTTCTGTT	112121
Qy	1080	tgtacttttggccttggatataaggtttaaataatgacatltgttgaacagcttcatltaa	1139
Db	112120	TGTACTTTTTGGGCTGGGACATGTGGTTTTCAAGGGAGACATCGTGTACTACACTTCATTTAA	112061

Qy	1140	aataacatatattgtcaaaa	1161
Db	112060	AATAACATATTTGTAAGAAA	112039

RESULT	13
AL135926	
LOCUS	158519 bp DNA PRI 30-NOV-2000
DEFINITION	Human DNA sequence from clone RP11-375F2 on chromosome 1 Contains a nucleotide similar to BR1 (ubiquitin-like 1 (sentrin)), a

pseudogene similar to ribosomal protein L29, ESTs, STSS and GSSs complete sequence.

ACCESSION	AL135926	GI:9801286
VERSION	AL135926.12	
KEYWORDS	HTG.	

SOURCE	ORGANISM
human.	human.
Homo sapiens	Homo sapiens
Eukaryota; Metazoa;	Eukaryota; Metazoa;
Chordata; Vertebrata;	Chordata; Vertebrata;
Euteleostomi	Euteleostomi

REFERENCE	AUTHORS	TAXA
1	Chapman, J.	Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.
		(bases 1 to 158519)

**TITLE** Direct Submission  
**JOURNAL** Submitted (03-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire  
**CR01 ISA** UK E-mail enquiries: humquerv@sanger.ac.uk Clone

requests: clonerequest@anger.ac.uk  
On Aug 14, 2000 this sequence version replaced gi:9621473.  
Duplicate sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the

validation annotation may not be representative of the validation corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequencing methods as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu), more care sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

EMBL: SwissPROT: TrEMBL: WORMPEP: Information

on the WormPeP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep). This sequence was obtained from a part of bacterial clone contigs of human

was generated from a bacterial artificial chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/chrom1>

<http://www.sanger.ac.uk/hnf/chr1>  
RP11-375F2 is from the library RPC1-11.2 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.utoronto.edu/>  
 VECTOR: pBAC3.6  
 IMPORTANT: This sequence is not the entire insert of clone

Rp11-375F2 may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone Rp11-375F2 is at 158519 in this

location/Qualifiers  
sequence. The true right end of clone RPI-10C1b is at 100 in this sequence.

```

1.158519
/organism="Homo sapiens"
/def="rayon.9606"

```

```

/clone="RP11-375F2"
/chromosome="1"

```

```
repeat_region
/c10ne_110_100_111
460..675
/note="MUTU repeat: matches 306..514 of consensus"
```

repeat_region	1403..1514	2 mer	tc	728 conserved"
repeat_region	1775..1824			

```

repeat_region      /note="MIR repeat: matches 65..133 of consensus
3415..3734         /note="L2 repeat: matches 2373..2707 of consensus"

```

```
repeat_region 4294.5519
```

1



QY 167 ttgacagagatagagatgagattcactcaaatgataatgacacacatctcaagaac 226  
 Db 31801 TTGGACAGAGATAGACAGGAGATTCACTCAAAAGTAAATGACAAACATCTCAAGAAC 31860  
 QY 227 tcaagaatcactatgataagaagaggtgttcaatgaaatcactcaaggcttctcttg 286  
 Db 31861 TCAAAAGATCACTACTGTCAAAGGAGGAGTTCAGATGAATTCAGATTGTCTTTG 31920  
 QY 287 agggcagagaatgctgataatcactcaagaacagcgaggaatggaggaagaatg 346  
 Db 31921 AAGATCAGAGATTCCTCTACTACTATTAAGAAAGAACTGGAGATGAGATGAAGATG 31980  
 QY 347 tgaatgaagattatagaagaagaagggggttcataacagcttagatattcttat 406  
 Db 31981 TGATTCAGATTTATCATGAGCAAA--TGAGGTCAATTCAGTATTAGTATTCCTTTAT 32038  
 QY 407 tttttctcttcctcaatcctcttcttcttcttcttcttcttcttcttcttcttct 466  
 Db 32039 TGTCTTTCTTTCCCAATCATTTTCTTTTAAATAGTCTTTCTTTCTTTCTTTCTTTG 32098  
 QY 467 ttcacaaaggaattgaaacatgacccacatctcttgaacatctgtaatttgaattc 526  
 Db 32099 CTCAAAGATGAAATGAAACTGGACCTCATCTCTTTAAACATCTGGTAAATTTGAATTC 32158  
 QY 527 tagtgcatactcaatctatgtgttcttctgctgctgctgctgctgctgctgctgctg 586  
 Db 32159 TAGTGCATATTCATTCATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 32218  
 QY 587 agtccctcaataatacctctctcttcttcttcttcttcttcttcttcttcttcttct 644  
 Db 32219 AGTCCCTCTGATTCCT 32278  
 QY 645 tttttgagacattgatttcttcttcttcttcttcttcttcttcttcttcttcttctt 704  
 Db 32279 TTTTGTAGGATTCATGATTTTTCAGGCTTTAGCAATTAAGATGAGCAAGCAAGT 32338  
 QY 705 ttcataatgacttcaaatgagccttgatgttc-agcatgtgatacttcttcttctgac 763  
 Db 32339 TTTGTAAAGACTTCCCAATGGCCCTAAAGTTCTAGCATGTATTCCTCTCTGAGC 32398  
 QY 764 tctgacttcaagtgagagatgagatttcttgaagaacagcagctgagaaatgactt 823  
 Db 32399 TGTACCTTTCAGTGGAGATGATGCTTTTCAGAGACTGAGACTGAGAAAAATGACTT 32458  
 QY 824 tcccttactggaagcacttcttcttcttcttcttcttcttcttcttcttcttcttctt 883  
 Db 32459 TCCCTTAATCTGAAGATAGCTTTAAAT--AGAGTCTAGACCAAAAGAGAGATAGCA 32516  
 QY 884 ggttgaagtcagaatgacagataagtgagagatgaatgacttcaagaatgagcttctc 943  
 Db 32517 TATTCAGATCAAGATGACATTAAGAGATGAGATGATGATGATGATGATGATGATG 32576  
 QY 944 tgaagaaagcatttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 993  
 Db 32577 TGAAGAAGATCATTTTAAGATCTAAATAATTAATAATCTTGTGAGAGATTCGA 32630

RESULT 14  
 AC011450/c AC011450 106168 bp DNA PRI 29-APR-2000  
 LOCUS Homo sapiens chromosome 19 clone CMC-30107, complete sequence.  
 DEFINITION AC011450  
 ACCESSION AC011450  
 VERSION AC011450.4 GI:7670120  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 106168)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 106168)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 REFERENCE 3 (bases 1 to 106168)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT On Apr 29, 2000 this sequence version replaced gi:7021594.  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.8% of Sequence:  
 Estimated Total Number of Errors is 0.3.  
 STS Content:  
 SHGC-34723 G28380  
 SHGC-58287 G38465  
 Location/Qualifiers  
 source  
 1..106168  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="19"  
 /clone="CTC-30107"  
 BASE COUNT 26156 a 24429 c 24340 g 31243 t  
 ORIGIN

Query Match 41.9%; Score 614.4; DB 87; Length 106168;  
 Best Local Similarity 73.0%; Pred. No. 6e-109;  
 Matches 1074; Conservative 0; Mismatches 71; Indels 327; Gaps 9;

QY 2 gggagagattgttaaaccggagagagaggttctgcttccggagcgctctgtgagga 61  
 Db 74682 GAGATGATTTTGGAACCTCGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 74623  
 QY 62 gacccgggtgagagacacgcctcatctctgacagagagcaaaccttcaactgag 121  
 Db 74622 GACCCGCGGTGAGGACGACGTCATCATGTGACGAGGCAAAACCTTCACCTGAGG 74563  
 QY 122 acttgaggatgaagagcaagtgatataataactcaaatgacttggacagagta 181  
 Db 74562 ACTTGGGGGATTAAGAAAGAGAGTGAATATATTAACCTCAAAATCACTGACAGGATAC 74503  
 QY 182 gtgagattcactcaaatgataaatgacacacatctcaagaacatcaaatcact 241  
 Db 74502 GTGAGATTCACCTTCAAGTAAATGACAAACATCTCAAGAACTCAAGAAATCATACT 74443  
 QY 242 gtcaagacagaggtgttccaatgataatcactcaggttctcttggaggtcagagaatg 301  
 Db 74442 GTCAAGACAGAGGATTCATGATGATCACTC-GGTTTCTTTGAGGGTCAAGAAATG 74384  
 QY 302 ctgataatcactcaaacgaagaactggaatggaagagagagatgtgattgattatc 361  
 Db 74383 CTGATTAATCATATCCCAAAAGAACTGGGAATGAGAGAAAGATGATGAGATTTATC 74324  
 QY 362 aggaacaaaggggggtcatcacaagcttagatatcttcttcttcttcttcttcttcc 421  
 Db 74324 AGGAACAAAGCTAGTGGGGGGGCTCATTTCA--ACGGCTTATCTTTTATTTTCC 74267  
 QY 422 tcaatcctt-----ttattttaaataagatcttcttgaatgtgtgttcaaaag 476  
 Db 74266 TCGATCTTTAATCTTTATTTTAAATAGTCTCTTTGTAATGTGGTGTGTAATAAGCC 74207  
 QY 477 aatgaaacatggaccccatctcttgaacacatctgtaattgaattctagctcat 536  
 Db 74206 AATGAAAACGTGATCCCAATCTTTGAAACATCTGTAAATTTGAATTTGGTGTTCAT 74147  
 QY 537 tattcattatgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 596  
 Db 74146 TATTCATATTCATTTGTTTCTTCTTCTGCTGATTTTGTGATCAAGGCTCAGTCCCTTC 74087

```

QY 597 atat----- 600
Db 74086 ANATTACCCCTCTTTT----- 74027
QY 601 ----- 600
Db 74026 TGTGTGGCGGAGTCTCGCTCACTGCAACCTCCACCTCCGGGGTTCATCTCATCTTCT 73967
QY 601 ----- 600
Db 73966 GCCTCAGCCTCCCGAGTAGCTGGATTACAGCAGCCGCCACAGCCTCGCTAATTTT 73907
QY 601 ----- 600
Db 73906 TGTATTTTCACTAGAGAGCGGCTTCAACCGTGTAGCCAGATGTCTCATCTCTGAT 73847
QY 601 ----- 600
Db 73846 CTGTGTATCTGCCAGCCTCGGCTCTTAAGTGTGGATTACAGCGCTGAGCCAGCCGCG 73787
QY 601 ----- 600
Db 73786 CCCAGCCTACCCCTCTCTTTTAAATTAAGTGTGACAGAGAGCGCAGCTTTTTCAGG 73727
QY 654 acattgcatcttcaagcct---gtgtgataaataagatcgacaatgcaatgctgata 710
Db 73726 ACATTGTGATTTTCAAGGCTTGTGTGTGATAGTAAAGATCGGCCAATGCAAGTGTTCATA 73667
QY 711 atgacttccaatgagcccgatgctc-aggatgctgatacttcaactctgagctgac 769
Db 73666 ATGACTTTTCAACTGCGCCCTGATGATGTCTAGCATGTGATGCTTCACTCTCGAGCTGTGAC 73607
QY 770 ttcaagtgagagatggaag-ttttcagagaactgagctgagaaatgacattcctt 828
Db 73606 TTTCAGTGGGAGATGGAGATTTTTCAGAGAACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 73547
QY 829 aacttgaagctacttttaaatgttgaggtctgagccaaagaagaatgacatgctg 888
Db 73546 ACCTTGAAGTACTTTTAAATTTGAGGCTGTGAGCCAAAGAGAGAAATATCAGTGTG 73487
QY 889 aagtcagaatgacagataagtgagagtaatgactaacccaaagaatgcttcaactgag 948
Db 73486 AAGTCAAGAGCAGACAGATGAGTGAAGTGTATGCTTAACCTCAAGTGTCTTCACTGAGG 73427
QY 949 aaaaagcattttaaagtttttaaaatctgtcagaagaatccagaagaatgcttaatt 1008
Db 73426 AAAAGCATTTTTAAAGATTTTAAATCTGTGCAAGATGCTCCAGAAAGTTCGATTT 73367
QY 1009 tcatagcaataataaagctatactacagagaatgaatacaagaacatgctctt 1068
Db 73366 TCATTATGCAATTATTAACATACAGTGAAGATGAATGAATGAAGAGCAACCTGCTCTT-- 73307
QY 1069 tgaatttcttctgacttttggcctgagataatggattttaaattgacatctgctgac 1128
Db 73308 ---TTTTATTTGTTCTTTTGGCTGTGAGATGAGGTTTAAATGACAT---TGTACC 73257
QY 1129 agcttcattaaataaacaatatgttcaaaa 1160
Db 73256 AGCTTCATTAAATAAACATATTTGTAATAA 73225

RESULT 15
AC068786 178733 bp DNA HTG 08-MAR-2001
LOCUS Homo sapiens chromosome 19 clone RP11-50111, WORKING DRAFT
DEFINITION
SEQUENCE, 15 unordered pieces.
ACCESSION AC068786.11 GI:13249431
VERSION AC068786.11
KEYWORDS HTG; HTGS_PHSF1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 178733)  
Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,  
Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,  
Mao, J., Lam, B., Marathe, R., Miranda, M., Morehouse, A.J., Nguyen, M.,  
Oelner, P., Palm, C.J., Ramirez, D., Southwick, A.M., Wilhelmy, J.,  
Yu, S., and Davis, R.W.  
Unpublished  
2 (bases 1 to 178733)  
Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,  
Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,  
Mao, J., Marathe, R., Morehouse, A.J., Oelner, P., Palm, C.J.,  
Ramirez, D., Wilhelmy, J., Yu, S., and Davis, R.W.  
Direct Submission  
Submitted (09-MAY-2000) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
On Mar 8, 2001 this sequence version replaced gi:13236646.  
-----  
Genome Center  
Center: Stanford DNA Sequencing and Technology Development  
Center  
Center code: SDSTDC  
Web site: <http://sequence-www.stanford.edu/group/human/>  
Contact: [hum-info@sequence.stanford.edu](mailto:hum-info@sequence.stanford.edu)  
-----  
Project Information  
Center project name: 899  
Center clone name: RP11-50111  
-----  
Summary Statistics  
Sequencing Vector: M13mp18; X02513  
Chemistry: Dye-primer; 0% of reads  
Assembly: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 168346 bases at least Q40  
Consensus quality: 171030 bases at least Q30  
Consensus quality: 172366 bases at least Q20  
Insert size: 178045; agarose-fp  
Insert size: 177333; sum-of-contigs  
Quality coverage: 9.2x in Q20 bases; agarose-fp  
Quality coverage: 9.2x in Q20 bases; sum-of-contigs.  
NOTE: This is a working draft sequence. It currently  
consists of 15 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
1  
1625: contig of 1625 bp in length  
1626  
1725: gap of unknown length  
1726  
3491: contig of 1766 bp in length  
3492  
3591: gap of unknown length  
3592  
5210: contig of 1619 bp in length  
5211  
5310: gap of unknown length  
5311  
7665: contig of 235 bp in length  
7666  
7765: gap of unknown length  
11506: contig of 3741 bp in length  
11507  
11606: gap of unknown length  
16778: contig of 5172 bp in length  
16779  
16878: gap of unknown length  
16879  
24169: contig of 7291 bp in length  
24170  
24269: gap of unknown length  
34333: contig of 10064 bp in length  
34334  
34433: gap of unknown length  
34434  
46447: contig of 12014 bp in length  
46448  
46547: gap of unknown length  
46548  
58761: contig of 12214 bp in length  
58762  
58861: gap of unknown length  
73753: contig of 14892 bp in length  
73754  
73853: gap of unknown length  
73854  
92184: contig of 18331 bp in length  
92185  
114445: gap of unknown length  
114446  
114445: contig of 22161 bp in length  
114446  
114545: gap of unknown length



Thu Aug 16 08:48:49 2001

us-09-484-964-1.rge

Page 20

QY 1129 agcttcattaaataaacacalatttgcacaa 1160  
|||||  
Db 121256 AGCTTCATTAAATAAACACATATTTCATAAAA 121287

Search completed: August 15, 2001, 16:06:02  
Job time: 3108 sec



model

Searched:

Total length: 0

Minimum	DB seq	Match
Maximum		0%
Minimum		100%

post-processor	Maximum existing first 45
----------------	---------------------------

Maximum first 45  
Listing first 45  
/geneseqn/NA1980.DAT:4  
1981.DAT:4

Database :

Listing -0601: \*

N.Generated	/gcgdata/geneseq/geneseq/NA1980.DAT *
1:	/SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT *
2:	/SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT *
3:	/SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT *
4:	/SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT *
5:	/SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT *
6:	/SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT *
7:	/SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT *
8:	/SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT *
9:	/SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT *
10:	/SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT *
11:	/SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT *
12:	/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT *
13:	/SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT *
14:	/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT *
15:	/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT *
16:	/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT *
17:	/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT *
18:	/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT *
19:	/SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT *
20:	/SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT *
21:	/SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT *
22:	/SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT *

... results predicted by chance to have a distribution.

pred. No. is greater than 0.07 and score greater by analysis of and is derived by analysis of SUMMARIES

Description

Result	Score	Query length	DB	ID
No				
1	1455	100.0	1455	19
2	1286.6	87.8	1514	21
3	1131.2	77.2	1420	19
4	1122.8	76.6	1196	19
5	469.2	32.0	616	21
6	368.8	25.2	372	18
7	368.8	25.2	372	18
8	345.2	23.6	425	17
9	200	13.7	936	21
10	163	11.1	936	22
C	11			

[illegible]

## ALIGNMENTS

Oligonucleotide D1  
Oligonucleotide D2  
Oligonucleotide D1  
Oligonucleotide D2  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D2  
Oligonucleotide D2  
Oligonucleotide D1  
Oligonucleotide D1  
Human gene thalita  
Arabidopsis thailia  
Arabidopsis thailia  
Arabidopsis thailia  
*zea mays* DNA fram  
plant microstelel  
plant microstele can  
human prostate can  
human secreted pro  
Plant microstelell  
Plant microstelell  
Human SWI3-1 rece  
human GABAR1 for Hu  
Genomic DNA thailia  
Arabidopsis can  
human prostate can  
Hemateopite D1  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide Exress  
Human brain fuence  
Expressed *Sesuvae*  
*Aspergillus oryzae*  
Plant microstelell  
Plant microstelell  
Plant microstelell  
*Aspergillus niger*

	RESULT	1	DNA:	1465 BP
ID	AAV34564			
XX	AAV34564	standard;		
AC	AAV34564	(first entry)		
XX	AAV34564;			
XX	11-SEP-1998	Homo sapiens sentrin-1 gene.		
XX	Homo sapiens	tumour neoplasia		
XX	sentrin-1; protection; tumour cell de-			
XX	Fas/APo-induced; detection; determination; ss-			
KW	detection;			
XX	Homo sapiens.			
OS				
XX	key	Location/Qualifi-		
FM	CDS	cation		
FT		/flag= a		
FT		/product= sen		
WT				
XX	<u>WO9820038-A1.</u>			
PN	14-MAY-1998.			
PD	05-NOV-1997;	97MO-US20344		
XX	05-NOV-1996;	9605-0030302		
PR	05-NOV-1996;			
PA	(TEXA ) UNIV TEXAS SYSTEM.			
PI	yeh ETH;			
DR	WPT: 1998-286868/25.			

Yeh Eih,  
WPI; 1998-286868/25

Thu Aug 16 08:48:50 2001

DR P-PSDB: AAM60079.

XX New isolated sentrin polypeptide(s) - which inhibit TNF receptor or  
PT Fas/APO-1-induced apoptosis, used to develop products for inducing  
PT cell death in tumours  
XX  
PS

Claim 10: Page 76-77; 120pp; English.  
The sequence is that encoding the sentrin-1 polypeptide.  
CC Sentrin polypeptides have the ability to protect or  
CC cells from tumour necrosis factor (TNF) or Fas/APO-1  
CC cell death (apoptosis) inhibitors of the sentrin polypeptide.  
CC Particularly, can be used for inducing cell death,  
CC determining the aggressiveness of a tumour and for detection and  
CC isolation of products. The products can also be used for  
XX detect a ubiquitin conjugating enzyme polypeptide or PML polypeptide.  
XX

Sequence 1465 BP; 455 A; 248 C; 292 G; 470 T; 0 other.

Query Match  
Best Local Similarity 100.0%; Score 1465; DB 19; Length 1465;  
Matches 1465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 cgggaagattgttaaacccggagcaggttcgtctaccgagccgctgctgagc 60
DB 1 cgggaagattgttaaacccggagcaggttcgtctaccgagccgctgctgagc 60
QY 61 agaccgccgggttaagccacccgagcaggttcgtctaccgagccgctgctgagc 60
DB 61 agaccgccgggttaagccacccgagcaggttcgtctaccgagccgctgctgagc 60
QY 121 gacttggggataaagaagaagcaggttcgtctaccgagccgagcagcactcaagc 120
DB 121 gacttggggataaagaagaagcaggttcgtctaccgagccgagcagcactcaagc 120
QY 181 agttagatcactcaagaatgaataatgaacacactcaagaagaactcgaagcagc 180
DB 181 agttagatcactcaagaatgaataatgaacacactcaagaagaactcgaagcagc 180
QY 241 tctcaagaagcaggttcgtctaccgagcaggttcgtctaccgagccgctgctgagc 240
DB 241 tctcaagaagcaggttcgtctaccgagcaggttcgtctaccgagccgctgctgagc 240
QY 301 gctgataatcactcaagaatgaataatgaacacactcaagaagaactcgaagcagc 300
DB 301 gctgataatcactcaagaatgaataatgaacacactcaagaagaactcgaagcagc 300
QY 361 caggaaacaaacgggggtcattcaacaggttcgaatgaagaagaactcgaagcagc 360
DB 361 caggaaacaaacgggggtcattcaacaggttcgaatgaagaagaactcgaagcagc 360
QY 421 ctaaatccctttatatttaataatagtccttgaatgggttcgaagaagaactc 420
DB 421 ctaaatccctttatatttaataatagtccttgaatgggttcgaagaagaactc 420
QY 481 gaaactgcacccacactcttgaacatcgtgaattgaattgaattgaattgaatt 480
DB 481 gaaactgcacccacactcttgaacatcgtgaattgaattgaattgaattgaatt 480
QY 541 catattgtttgtttcatctgtcgtattgttgatcaagcagcagcagcagcagc 540
DB 541 catattgtttgtttcatctgtcgtattgttgatcaagcagcagcagcagcagc 540
QY 601 taacctcccttttaataatagtcgtgaattgttgatcaagcagcagcagcagc 600
DB 601 taacctcccttttaataatagtcgtgaattgttgatcaagcagcagcagcagc 600
QY 661 attcagagcttggttgataaataagatcgaagaatggttcgaagaagaactcgc 660
DB 661 attcagagcttggttgataaataagatcgaagaatggttcgaagaagaactcgc 660
QY 720 atttcaagccttggttgataaataagatcgaagaatggttcgaagaagaactcgc 720
DB 720 atttcaagccttggttgataaataagatcgaagaatggttcgaagaagaactcgc 720
```

us-09-484-964-1.rng

Page 2

```
QY 721 atttgccctgattgacagcaggttacttacttacttacttacttacttacttactt 720
DB 721 atttgccctgattgacagcaggttacttacttacttacttacttacttacttactt 720
QY 781 gttgaaatttttcaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 780
DB 781 gttgaaatttttcaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 780
QY 841 cttaaaatttttgggggttcgagcagcagcagcagcagcagcagcagcagcagcagc 840
DB 841 cttaaaatttttgggggttcgagcagcagcagcagcagcagcagcagcagcagcagc 840
QY 901 cagataagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900
DB 901 cagataagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900
QY 961 aagatttttaaaatcttgcagaagaatccagaagaatccagaagaatccagaagaat 960
DB 961 aagatttttaaaatcttgcagaagaatccagaagaatccagaagaatccagaagaat 960
QY 1021 aataagctatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1020
DB 1021 aataagctatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1020
QY 1081 gtaacttttgccctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080
DB 1081 gtaacttttgccctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080
QY 1141 ataaatatttttcaaaatcgtactaataatgacttacttacttacttacttacttact 1140
DB 1141 ataaatatttttcaaaatcgtactaataatgacttacttacttacttacttacttact 1140
QY 1201 aaaaatgcttaaaatgacttacttacttacttacttacttacttacttacttactt 1200
DB 1201 aaaaatgcttaaaatgacttacttacttacttacttacttacttacttacttactt 1200
QY 1261 ttttcttacttacttacttacttacttacttacttacttacttacttacttacttact 1260
DB 1261 ttttcttacttacttacttacttacttacttacttacttacttacttacttacttact 1260
QY 1321 agtgccttctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320
DB 1321 agtgccttctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320
QY 1381 ttttcttacttacttacttacttacttacttacttacttacttacttacttacttact 1380
DB 1381 ttttcttacttacttacttacttacttacttacttacttacttacttacttacttact 1380
QY 1441 ttttgaataatggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1440
DB 1441 ttttgaataatggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1440
```

RESULT 2  
ID AA235859 standard; cDNA, 1514 BP.  
AC AA235859;  
DT 03-FEB-2000 (first entry)  
XX Human sentrin nucleotide sequence.  
XX Human; sentrin; antisense oligonucleotide; phosphorothioate;  
XX Inhibition; modulation; expression; diagnosis; ss.  
OS Homo sapiens.  
Key  
CDS Location/Qualifiers  
136..441  
/\*tag= a  
/product= "sentrin"

*antiser*

PN US5985664-A.  
 XX 16-NOV-1999.  
 XX 17-DEC-1998; 98US-0213768.  
 XX 17-DEC-1998; 98US-0213768.  
 XX (ISIS-) ISIS PHARM INC.  
 XX Baker BF, Cowsett LM;  
 PI WPI: 2000-022284/02.  
 DR P-PSDB: AAY49967.  
 XX  
 PT Antisense compound which modulates human sentrin expression, useful for  
 treating diseases associated with sentrin expression -  
 XX  
 PS Example 13: Column 39-42; 29pp; English.  
 XX  
 CC The present invention describes an antisense compound (I) 8-30  
 CC nucleotides long targeted to a nucleic acid molecule encoding human  
 CC sentrin. The antisense compound comprises a phosphorothioate antisense  
 CC oligonucleotide which inhibits expression of human sentrin. (I) is  
 CC useful for inhibiting expression of sentrin in human cells or tissues  
 CC in vitro, for treating humans or other animals suspected of having or  
 CC being prone to a disease associated with sentrin expression. (I) can  
 CC also be used for research or diagnostic purposes. The present  
 CC sequence encodes human sentrin.  
 XX  
 SQ Sequence 1514 BP; 462 A; 256 C; 314 G; 482 T; 0 other;

Query Match 87.8%; Score 1286.6; DB 21; Length 1514;  
 Best Local Similarity 94.1%; Pred. No. 1.9e-278;  
 Matches 1384; Conservative 0; Mismatches 74; Indels 13; Gaps 4;

QY 2 ggggaagattgttaaaccccgagagaggtctgtctaccgagcgctgtctgtcga 61  
 DB 50 ggggaagattgttaaaccccgagagaggtctgtctaccgagcgctgtctgtcga 109  
 QY 62 gaccccgaggtgagccacgtcatatgtctgacagagaggaacccctcaacgag 121  
 DB 110 gaccccgaggtgagccacgtcatatgtctgacagagaggaacccctcaacgag 169  
 QY 122 acttgaggagtaagaaagagtgatataataactcaagtcatttgacagagataga 181  
 DB 170 acttgaggagtaagaaagagtgatataataactcaagtcatttgacagagataga 229  
 QY 182 gtgagatcactcaaaagtgaaatgacacacatctcaagaacccaagaatcact 241  
 DB 230 gtgagatcactcaaaagtgaaatgacacacatctcaagaacccaagaatcact 289  
 QY 242 gtcaaaagcaggggtgtccatgaatcactcaggttctcttgagggtcagaagt 301  
 DB 290 gtcaaaagcaggggtgtccatgaatcactcaggttctcttgagggtcagaagt 349  
 QY 302 ctgataactcactcaaaagtaactggaaatggaggaagaagtgtgagtgaattatc 361  
 DB 350 ctgataactcactcaaaagtaactggaaatggaggaagaagtgtgagtgaattatc 409  
 QY 362 aggaacaaaggggggtgtcattcaacagtttagatatctttttttttttttttcc 421  
 DB 410 aggaacaaaggggggtgtcattcaacagtttagatatctttttttttttttttcc 469  
 QY 422 tcaatcctttcc 481  
 DB 470 tcaatcctttcc 529  
 QY 482 aaactggaccccatctcttgaacaatctgtgaattgaaatctgtctcatattc 541  
 DB 530 aaactggaccccatctcttgaacaatctgtgaattgaaatctgtctcatattc 589

QY 542 attatgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 601  
 DB 590 attatgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 649  
 QY 602 accctccctttcc 661  
 DB 650 accctccctttcc 709  
 QY 662 ttccaagctgtgtgtgataaataagatcgacaatgcaagtgttcaataatgtcttcca 721  
 DB 710 ttccaagctgtgtgtgataaataagatcgacaatgcaagtgttcaataatgtcttcca 764  
 QY 722 attgacctgtgtgtgataaataagatcgacaatgcaagtgttcaataatgtcttcca 781  
 DB 765 tggagagatgaaagttttgtgtgtgataaataagatcgacaatgcaagtgttcaata 824  
 QY 782 atggaagttt-----tcagagactgaaactgtggaanaatgaccccttccaattga 835  
 DB 825 ctcccaattgacctgacagagagactgaaactgtggaanaatgaccccttccaattga 884  
 QY 836 agctactttaaa-tttgaggtctgtgacccaagaagaaggaatcaggttgaagtca 894  
 DB 885 agctactttaaa-tttgaggtctgtgacccaagaagaaggaatcaggttgaagtca 944  
 QY 895 agatgacagataaggtgtgagtaataactcacaagatggtctcaactgaaagaaag 954  
 DB 945 agatgacagataaggtgtgagtaataactcacaagatggtctcaactgaaagaaag 1004  
 QY 955 catttaagattttttaaaatctgttcagaagaatccagaanaagttcatttcaata 1014  
 DB 1005 catttaagattttttaaaatctgttcagaagaatccagaanaagttcatttcaata 1064  
 QY 1015 gcaatataaagctatcagcagaaatgataacaaagaacactgctcttttaagt 1074  
 DB 1065 gcaatataaagctatcagcagaaatgataacaaagaacactgctcttttaagt 1123  
 QY 1075 ttattgtacttttggcctgtgagataagttttaaagacatgtgtcgtacagcttc 1134  
 DB 1124 ttattgtacttttggcctgtgagataagttttaaagacatgtgtcgtacagcttc 1183  
 QY 1135 attaaataaacaataattgttcaanaacgtactaaatgtctatttttaattgtata 1194  
 DB 1184 attaaataaacaataattgttcaanaacgtactaaatgtctatttttaattgtata 1243  
 QY 1195 gaaagaaaaaattgctaaataaagtttctgtcataaactggaattgacatagt 1254  
 DB 1244 gaaagaaaaaattgctaaataaagtttctgtcataaactggaattgacatagt 1303  
 QY 1255 acaaattttctcattactgttacaggatgtgttaatgacttggagcactgaaagt 1314  
 DB 1304 acaaattttctcattactgttacaggatgtgttaatgacttggagcactgaaagt 1363  
 QY 1315 tactgaagtgtcctctgaaatcaagaatttaataaggccacaataccttttaatacca 1374  
 DB 1364 tactgaagtgtcctctgaaatcaagaatttaataaggccacaataccttttaatacca 1423  
 QY 1375 ggtgtgtgtttcc 1434  
 DB 1424 ggtgtgtgtttcc 1483  
 QY 1435 caatcatgttgataaattggcagtcagcc 1465  
 DB 1484 caatcatgttgataaattggcagtcagcc 1514

RESULT 3  
 AAV59557  
 ID AAV59557 standard; DNA; 1220 BP.

XX AAV59557;  
 AC  
 XX 06-JAN-1999 (first entry)  
 DT  
 XX

DE	Human secreted protein gene 47 clone HOGAN7.5.
XX	
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS	
XX	Homo sapiens.
PN	
XX	WO9839448-A2.
PD	
XX	
XX	11-SEP-1998.
XX	
XX	
XX	06-MAR-1998; 98WO-US04493.
XX	
XX	
XX	02-OCT-1997; 97US-0061060.
PR	07-MAR-1997; 97US-0038621.
PR	07-MAR-1997; 97US-0040161.
PR	07-MAR-1997; 97US-0040162.
PR	07-MAR-1997; 97US-0040163.
PR	07-MAR-1997; 97US-0040333.
PR	07-MAR-1997; 97US-0040334.
PR	07-MAR-1997; 97US-0040336.
PR	07-MAR-1997; 97US-0040626.
PR	11-APR-1997; 97US-0043311.
PR	11-APR-1997; 97US-0043312.
PR	11-APR-1997; 97US-0043313.
PR	11-APR-1997; 97US-0043314.
PR	11-APR-1997; 97US-0043568.
PR	11-APR-1997; 97US-0043569.
PR	11-APR-1997; 97US-0043576.
PR	11-APR-1997; 97US-0043578.
PR	11-APR-1997; 97US-0043580.
PR	11-APR-1997; 97US-0043669.
PR	11-APR-1997; 97US-0043670.
PR	11-APR-1997; 97US-0043671.
PR	11-APR-1997; 97US-0043672.
PR	11-APR-1997; 97US-0043674.
PR	23-MAY-1997; 97US-0047492.
PR	23-MAY-1997; 97US-0047500.
PR	23-MAY-1997; 97US-0047501.
PR	23-MAY-1997; 97US-0047502.
PR	23-MAY-1997; 97US-0047503.
PR	23-MAY-1997; 97US-0047581.
PR	23-MAY-1997; 97US-0047582.
PR	23-MAY-1997; 97US-0047583.
PR	23-MAY-1997; 97US-0047584.
PR	23-MAY-1997; 97US-0047585.
PR	23-MAY-1997; 97US-0047586.
PR	23-MAY-1997; 97US-0047587.
PR	23-MAY-1997; 97US-0047588.
PR	23-MAY-1997; 97US-0047589.
PR	23-MAY-1997; 97US-0047590.
PR	23-MAY-1997; 97US-0047592.
PR	23-MAY-1997; 97US-0047593.
PR	23-MAY-1997; 97US-0047594.
PR	23-MAY-1997; 97US-0047595.
PR	23-MAY-1997; 97US-0047597.
PR	23-MAY-1997; 97US-0047598.
PR	23-MAY-1997; 97US-0047599.
PR	23-MAY-1997; 97US-0047600.
PR	23-MAY-1997; 97US-0047601.
PR	23-MAY-1997; 97US-0047612.
PR	23-MAY-1997; 97US-0047613.
PR	23-MAY-1997; 97US-0047614.
PR	23-MAY-1997; 97US-0047615.
PR	23-MAY-1997; 97US-0047617.
PR	23-MAY-1997; 97US-0047618.

sample or by determining the presence of mutations in the new  
CC polynucleotides. Specific uses are described for each of the 186  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AAV9511 for described uses).

XX Sequence 1220 BP; 382 A; 207 C; 258 G; 372 T; 1 other;

Query Match 77.2%; Score 1131.2; DB 19; Length 1220;  
Best Local Similarity 99.6%; Pred. No. 9,7e-244;  
Matches 1153; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

OY 2 gggaaagatttgtaaaccccgagcagaggtctctcttaaccggagcgctgctgctgga 61  
DB 42 gagaagatttgtaaaccccgagcagaggtctctcttaaccggagcgctgctgctgga 101  
OY 62 gaccccggtgaagcccgatcatgctctgacacgaagcgaacactcaactgag 121  
DB 102 gaccccggtgaagcccgatcatgctctgacacgaagcgaacactcaactgag 161  
OY 122 actgggggataagaagcaggtgataatataaactcaagtcattgagcagatgca 181  
DB 162 actgggggataagaagcaggtgataatataaactcaagtcattgagcagatgca 221  
OY 182 gtgagattacttcaagaagtaaatgacacacatctcaagaactcaagaatcact 241  
DB 222 gtgagattacttcaagaagtaaatgacacacatctcaagaactcaagaatcact 281  
OY 242 gtcaaaagcaggtgtctcaatgtaaltcactcaggttctctcttgaaggtcagaatg 301  
DB 282 gtcaaaagcaggtgtctcaatgtaaltcactcaggttctctcttgaaggtcagaatg 341  
OY 302 ctgataatcaactccaagaagactggaatggaagaagaatgagtgatgagttatc 361  
DB 342 ctgataatcaactccaagaagactggaatggaagaagaatgagtgatgagttatc 401  
OY 362 aggaacaacgggggtgcatcacaagttagatattcttattttttttttccc 421  
DB 402 aggaacaacgggggtgcatcacaagttagatattcttattttttttttccc 461  
OY 422 tcaatccttttatt 481  
DB 462 tcaatccttttatt 521  
OY 482 aaactgaccccatctcttgaacatctgtaattgtaattgtaattgtaattgtaatt 541  
DB 522 aaactgaccccatctcttgaacatctgtaattgtaattgtaattgtaattgtaatt 581  
OY 542 attattgttt 601  
DB 582 attattgttt 641  
OY 602 accctctcttt 661  
DB 642 accctctcttt 701  
OY 662 ttttcaggtctgtgtgtaataaagatcgacacgaaggttcccttttcaggaattgca 721  
DB 702 ttttcaggtctgtgtgtaataaagatcgacacgaaggttcccttttcaggaattgca 761  
OY 722 atggccctgagttc-agcatgattacttactcctctgacgtgacttccagtgga 780  
DB 762 atggccctgagttc-agcatgattacttactcctctgacgtgacttccagtgga 821  
OY 781 gatggaagttttcagagacgtgaactgtgaaaaaagaccttcccttaacttgaagta 840  
DB 822 gatggaagttttcagagacgtgaactgtgaaaaaagaccttcccttaacttgaagta 881  
OY 841 cttttaaatttgaaggtctggaacaaagaagaagaatataagttgaagtaagta 900  
DB 882 cttttaaatttgaaggtctggaacaaagaagaagaatataagttgaagtaagta 941  
OY 901 cagataagtgagagataactcaagaatggtcttcaactgaagaagaagcatttt 960

DB 942 cagataagtgagagataactcaagaatggtcttcaactgaagaagaagcatttt 1001  
OY 961 aagattttttaaacttttcaagaagatcccgaaagttcaattttcattgaacatt 1020  
DB 1002 aagattttttaaacttttcaagaagatcccgaaagttcaattttcattgaacatt 1061  
OY 1021 aataagctatactgacagaatgataacacagaaacacgtctttttagattttatt 1080  
DB 1062 aataagctatactgacagaatgataacacagaaacacgtctctttt-gattttatt 1120  
OY 1081 gtaacttttggcctggaggtatgggttttaaatgacattgtctgatacagattcaaa 1140  
DB 1121 gtaacttttggcctggaggtatgggttttaaatgacattgtctgatacagattcaaa 1180  
OY 1141 ataacaatatgtgcaaaa 1160  
DB 1181 ataacaatatgtgcaaaa 1200

## RESULT 4

AAV59722 standard; DNA; 1196 BP.

AAV59722;

19-JAN-1999 (first entry)

Human secreted protein gene 47 clone HOGAV75.

Human; secreted protein; fusion protein; gene therapy; protein therapy;  
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

MO9839448-A2.

11-SEP-1998.

06-MAR-1998; 98MO-US04493.

02-OCT-1997; 97US-0061060.  
07-MAR-1997; 97US-0038621.  
07-MAR-1997; 97US-0040161.  
07-MAR-1997; 97US-0040162.  
07-MAR-1997; 97US-0040163.  
07-MAR-1997; 97US-0040333.  
07-MAR-1997; 97US-0040334.  
07-MAR-1997; 97US-0040336.  
07-MAR-1997; 97US-0040336.  
07-MAR-1997; 97US-0043626.  
11-APR-1997; 97US-0043311.  
11-APR-1997; 97US-0043312.  
11-APR-1997; 97US-0043313.  
11-APR-1997; 97US-0043314.  
11-APR-1997; 97US-0043568.  
11-APR-1997; 97US-0043569.  
11-APR-1997; 97US-0043576.  
11-APR-1997; 97US-0043578.  
11-APR-1997; 97US-0043580.  
11-APR-1997; 97US-0043580.  
11-APR-1997; 97US-0043669.  
11-APR-1997; 97US-0043670.  
11-APR-1997; 97US-0043671.  
11-APR-1997; 97US-0043672.  
11-APR-1997; 97US-0043674.  
11-APR-1997; 97US-0043674.  
23-MAY-1997; 97US-0047500.  
23-MAY-1997; 97US-0047500.  
23-MAY-1997; 97US-0047501.

[illegible]

```

DB 528 aaaaacggaccacccaccccttgaacacggaatctgaattcagtcgacattc 587
QY 542 attatgtgtgtttcatatgtgtcatttttggatgaacccctcagccctcatt 601
DB 588 attatgtgtgtttcatatgtgtcatttttggatgaacccctcagccctcatt 647
QY 602 accctcctcttttaaaataacgtgtgcacagagaggtcaccttttcagagatgca 661
DB 648 accctcctcttttaaaataacgtgtgcacagagaggtcaccttttcagagatgca 707
QY 662 ttttaggtgtgtgttgaataaagatcgacacgaatgaaggttcataaagatttca 721
DB 708 ttctcaggtgtgtgtgaataaagatcgacacgaatgaaggttcataaagatttca 767
QY 722 attgcccctgaagttc-agacgtatattactcactcctgagctgtgacttcagtgga 780
DB 768 attgcccctgaagttcagatgtatctactcctgagctgtgacttcagtgga 827
QY 781 gatggaagttttcagagaactgaactgtggaataatgaccttccctaacttgaagcta 840
DB 828 gatggaagttttcagagaactgaactgtggaataatgaccttccctaacttgaagcta 887
QY 841 ctttaaaatttgaggtgtgcacaaagaagaggaataataggtttgaagtcagatga 900
DB 888 ctttaaaatttgaggtgtgcacaaagaagaggaataataggtttgaagtcagatga 947
QY 901 cagataagtgagagtaagtaacttaactcacaagaatgtctcactgaagaagaagattt 960
DB 948 cagataagtgagagtaagtaacttaactcacaagaatgtctcactgaagaagaagattt 1007
QY 961 aagattttttaaataatctgtcagaagaatcccaagaagaattcattcattcagaatt 1020
DB 1008 aagattttttaaataatctgtcagaagaatcccaagaagaattcattcattcagaatt 1067
QY 1021 aataaagctatatactgcagaataatgaatacaacagaacactgtcttttaagtattt 1080
DB 1068 aataaagctatatactgcagaataatgaatacaacagaacactgtctttt-gatttattt 1126
QY 1081 gtacttttggcctggagatgtgttttaaatgagacatgtctgtacacagctcattaa 1140
DB 1127 gtacttttggcctggagatgtgttttaaatgagacatgtctgtacacagctcattaa 1186
QY 1141 ataacaata 1150
DB 1187 ataacaata 1196

```

## RESULT 5

AAAI6215 standard; DNA; 616 BP.

AC AAAI6215;

DT 14-JUN-2000 (first entry)

DE Human colon cancer differentially expressed nucleotide sequence #220.

XX Colon cancer; detect; differential expression; human; treatment;

KM detect mutation; non-invasive diagnostic method; ds.

OS Homo sapiens.

PN WO200012702-A2.

PD 09-MAR-2000.

PF 30-AUG-1999; 99MO-US19424.

PR 31-AUG-1998; 98US-0098639.

XX 27-JAN-1999; 99US-0117393.

PA (FARB ) BAYER CORP.

```

XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
PI Catino TJ, D'Wivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
PI Schlegel R;
XX WPI: 2000-256641/22.
PT Novel nucleic acids and proteins for identifying therapeutic agents
PT useful for treating and diagnosing cancer, especially colon cancer
PS Claim 16; Page 211-212; 345pp; English.
XX
CC This sequence represents a human nucleotide sequence which is
CC differentially expressed in colon cancer cells compared to the expression
CC levels in normal cells. The nucleotide sequence can be used as a source
CC of primers and probes. The nucleotide sequence is useful for determining
CC the phenotype of a cell by detecting the differential expression of the
CC sequence relative to a normal cell. The probes derived from the sequence
CC can also be used to determine the phenotype of cells in a sample. Probes
CC and antibodies which hybridise to the nucleotide sequence can also be
CC used to determine the phenotype of a cell. The primers are useful for
CC detecting a mutation in a test nucleotide sequence and also for detecting
CC cancer, preferably colon cancer. Antibodies against the protein encoded
CC by the nucleotide sequence can also be used in a method to detect colon
CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
CC colon cancer at an early stage.
XX
S0 Sequence 616 BP; 181 A; 109 C; 149 G; 170 T; 7 other:

```

## Query Match

32.0%; Score 469.2; DB 21; Length 616;

Best Local Similarity 95.2%; Pred. No. 7; 1e-96;

Matches 553; Conservative 0; Mismatches 20; Indels 8; Gaps 7;

```

QY 2 ggaaggaattgttaaaccccgagagaggtctgtcttaaccgagcgctgtgtcgga 61
DB 41 gagaaggaattgttaaaccccgagagaggtctgtcttaaccgagcgctgtgtcgga 100
QY 62 gaaccgaggtgaagcaccgcatcatatgtctgacacagagagcacaacattcagag 121
DB 101 gaaccgaggtgaagcaccgcatcatatgtctgacacagagagcacaacattcagag 160
QY 122 actgggggataagaagaaggtgataatatttaactcaagtcattgacagataga 181
DB 161 actgggggataagaagaaggtgataatatttaactcaagtcattgacagataga 220
QY 182 gtgagattcacttcaaatgaaatgacacacatctcaagaactcaagaatcact 241
DB 221 gtgagattcacttcaaatgaaatgacacacatctcaagaactcaagaatcact 280
QY 242 gtcaagaagcaggtgttccaatgaattcactcaggtctcttggagggtcagaattg 301
DB 281 gtcaagaagcaggtgttccaatgaattcactcaggtctcttggagggtcagaattg 340
QY 302 ctgataatcatactccaaagaactggaatgaggaagaagtgtg-aaagtat 360
DB 341 ctgataatcatactccaaagaactggaatgaggaagaagtgtg-aaagtat 400
QY 361 cagaacaacagggggtcatcacaagttagatattcattttttttttcc 420
DB 401 cagaacaacagggggtcatcacaagttagatattcattttttttttttc 459
QY 421 ctcaatccttttttttttttttttttttttttttttttttttttttttttttttt 480
DB 460 ctcaatccttttttttttttttttttttttttttttttttttttttttttttttt 518
QY 481 gaaactggcaccacatctcttgaacaactgtgaattgaattcagttcattat 540
DB 519 gaaactggcaccacatctcttgaacaactgtgaattgaattcagttcattat 575
QY 541 cattatgtgttttcatatgtgtcagattttgtgtgacagc 583
DB 576 cattatgtgtgtt--cattgcmgaattttgggagtaaac 616

```

## RESULT 6

AAT63339 standard; CDNA: 372 BP.

AAT63339;

17-AUG-1997 (first entry)

Human host cell protein NP1-6 partial cDNA clone.

Host cell protein; NP1-6; nucleoprotein interactor 6;  
influenza virus; replication; antiviral; virucide; ss.

Homo sapiens.

MO9712967-A1.

10-APR-1997.

06-OCT-1995; 95WO-US13044.

06-OCT-1995; 95WO-US13044.

(MOUN ) MOUNT SINAI MEDICAL CENT.

Oneill R, Palese P;

WPI; 1997-226211/20.

New isolated DNA which encodes viral interacting proteins - used in  
assays to isolate products for inhibiting viral protein binding  
which is required for infection, replication, assembly or release

Disclosure; Fig 11; 98pp; English.

CDNA clones (AAT63335-39) comprise partial sequences for human  
nucleoprotein interactor proteins NP1-2 to NP1-6, respectively.  
These are host cell proteins which interact with influenza virus  
nucleoprotein (NP) and which may be accessory proteins required for  
influenza virus replication. NP1-6 was identified as a novel  
protein sequence. NP1 sequences were isolated by interactive trap  
selection using LexA-NP as bait and yeast transformed with an Hela  
cDNA library. NP1-1 (see also AAT63334) and NP1-2 to NP1-6 cDNAs were  
cloned and analysed. The clones can be used to produce human NP1  
proteins for use in identifying cpts. that inhibit viral  
replication.

Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

Query Match 25.2%; Score 368.8; DB 18; Length 372;

Best local Similarity 99.5%; Pred. No. 1.7e-73; Mismatches 2; Indels 0; Gaps 0;

Matches 370; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

9 attgttaaccccgagcggtctgtcttaccgagcgccgtctgtgagagacccc 68

1 attgttaaccccgagcggtctgtcttaccgagcgccgtctgtgagagacccc 60

69 ggggtgaagcaccgctcatctgtctgacgagggcaaacctcaactgaagacttg 128

61 ggggtgaagcaccgctcatctgtctgacgagggcaaacctcaactgaagacttg 120

129 ggtatgaagcaccgctcatctgtctgacgagggcaaacctcaactgaagacttg 188

121 ggtatgaagcaccgctcatctgtctgacgagggcaaacctcaactgaagacttg 180

189 tcactcaaaagtgaatgaacaacacatctcaagaactcaagaatcacactgcaag 248

181 tcactcaaaagtgaatgaacaacacatctcaagaactcaagaatcacactgcaag 240

249 acaggggtgttccatgaatcactcaggttctcttgaaggtcagagaatgtctgataa 308

## RESULT 7

AAF59399 standard; DNA: 372 BP.

AAF59399;

02-MAY-2001 (first entry)

Human host cell protein NP1-6 partial nucleotide sequence SEQ ID NO:11.

Identification; antiviral; viral protein; viral replication; NP;  
viral infection; nucleoprotein; ds.

Homo sapiens.

WO200111335-A2.

15-FEB-2001.

11-AUG-2000; 2000WO-US22257.

11-AUG-1999; 99US-0148263.

(MOUN ) MOUNT SINAI SCHOOL MEDICINE.

O'Neill R, Harty R, Palese PM;

WPI; 2001-168816/17.

Identifying a substance that inhibits the interaction between a viral  
protein and a host cell protein, useful for the discovery of new  
antiviral compounds -

Disclosure; Fig 11; 147pp; English.

The present invention describes a method (M1) for identifying a  
substance that inhibits the interaction of a viral protein (VP) with a  
host cell protein (HP). The method comprises: (a) contacting HP with VP  
in the presence of a test substance; and (b) detecting complex formation,  
where the ability of the test substance to inhibit HP/VP interaction is  
indicated by a decrease in complex formation. The antiviral compounds  
that inhibit the interaction between a host protein (NS1-BP or NP1-1)  
and a viral protein (NS1) are useful for treating or inhibiting viral  
infection, preferably influenza and rhadovirus infection, in humans.

Antiviral compounds include peptides and antibodies. In particular  
compositions comprising a polypeptide containing an amino acid sequence  
corresponding to the NP-NS1 domain of the influenza virus NP protein,  
which inhibits the specific interaction of the NP1-1 protein with the  
influenza virus NP protein are useful for treating or inhibiting  
influenza viral infection in humans. The present sequence represents a  
human host cell protein NP1-6 partial nucleotide sequence, which is used  
in the exemplification of the present invention.

Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

Query Match 25.2%; Score 368.8; DB 22; Length 372;

Best local Similarity 99.5%; Pred. No. 1.7e-73; Mismatches 2; Indels 0; Gaps 0;

Matches 370; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

9 attgttaaccccgagcggtctgtcttaccgagcgccgtctgtgagagacccc 68



```

Db      1 atttgtaaaccccgagagaggtctctgttaaccgagcgctgtgaggaacccc 60
Oy      69 ggggtgaagccacgcgtcatcatcgtctgaccagagagcaaaccttaactgagacttgg 128
Db      61 ggggtgaagccacgcgtcatcatcgtctgaccagagagcaaaccttaactgagacttgg 120
Oy      129 ggataagaagcaggtgatatataataaactcaagtcattgacagagatagcagttagat 188
Db      121 ggataagaagcaggtgatatataataaactcaagtcattgacagagatagcagttagat 180
Oy      189 tcaactcaaatgtaaatagacaacacacatcaagaactcaagaatatactcttcaag 248
Db      181 tcaactcaaatgtaaatagacaacacacatcaagaactcaagaatatactcttcaag 240
Oy      249 acaagggtgtccaatgaattcactcaggtttctcttgaagggtcagagaatttgcataa 308
Db      241 acaagggtgtccaatgaattcactcaggtttctcttgaagggtcagagaatttgcataa 300
Oy      309 tcaactccaaaagagactgggaatgagagaagatgtgattgaaattatcaggaa 368
Db      301 tcaactccaaaagagactgggaatgagagaagatgtgattgaaattatcaggaa 360
Oy      369 aacggggggtca 380
Db      361 aacggggggtca 372

```

RESULT 8  
AAT03735  
ID AAT03735 standard; cDNA: 425 BP.

AC AAT03735;

DT 26-MAR-1996 (first entry)

DE TNF-R p55IC/Fas-IC-binding protein DD11 cDNA clone.

KW Tumour necrosis factor receptor; TNF-R; p55IC; Fas-IC;  
KW Intracellular domain binding protein; ss.

OS Homo sapiens.

PN W09531544-A1.

PD 23-NOV-1995.

PF 11-MAY-1995; 95WO-US05854.

PR 02-OCT-1994; 94IL-0111125.

PR 11-MAY-1994; 94IL-0109632.

PA (WEIN/) WEINMURZEL H.  
PA (YEDA) YEDA RES & DEV CO LTD.

PI Boldin M, Mett I, Varfolomeev E, Wallach D;

DR WPI; 1996-010930/01.

PT TNF-NGF receptor superfamily intracellular domain-binding proteins -  
PT useful for modulating receptor function, e.g. for treating tumours  
PT or HIV-infected cells

PS Claim 18; Fig 12; 96pp; English.

CC cDNA clone DD11 (AAT03735) was isolated by screening a human HeLa  
CC library using the human tumour necrosis factor receptor (TNF-R)  
CC p55 death domain (p55DD) as 'bait'. The full length of the DD11  
CC transcript is approx 1.2 kb. The protein encoded by cDNA DD11  
CC interacts strongly with p55DD (amino acids 326-414) and also with  
CC human and mouse FAS intracellular domain. Such proteins may  
CC be used modulate TNF-R function.

Sequence 425 BP; 135 A; 80 C; 115 G; 95 T; 0 other;

Query Match 23.6%; Score 345.2; DB 17; Length 425;  
Best Local Similarity 95.7%; Pred. No. 3.3e-68;  
Matches 377; Conservative 0; Mismatches 13; Indels 4; Gaps 2;

```

Oy      2 ggggaagatttgaataacccggagcgaggttctgttaccggcgccgtgtgaggga 61
Db      32 gagaagatttgaataacccggagcgaggttctgttaccggcgccgtgtgaggga 91
Oy      62 gaccccggttgaagccacgcatcatcgtctgaccagagagcaaacctt-caacttgg 120
Db      92 gaccccggttgaagccacgcatcatcgtctgaccagagagcaaaccttcaacttgg 151
Oy      121 gacttggggataaagaagcaggtgaaatataataaactcaagaatcttgacagatagc 180
Db      152 gacttggggataaagaagcaggtgaaatataataaactcaagaatcttgacagatagc 211
Oy      181 agtgaattcactcaagtgaaatgacaacacatctcaagaactcaagaatcatatgc 240
Db      212 agtgaattcactcaagtgaaatgacaacacatctcaagaactcaagaatcatatgc 271
Oy      241 tgtcaagaacaggggttccaatgaattcactcaggtttctcttgaagggtcagagaatt 300
Db      272 tgtcaagaacaggggttccaatgaattcactcaggtttctcttgaagggtcagagaatt 331
Oy      301 gctgataatcactaccacaagaactggaatgagagaaga---tgtgattgaagt 357
Db      332 gctgataatcactaccacaagaactggaatgagagaagaatgattgattgaagt 391
Oy      358 tatcagaacaacaggggggtcattcaacagtt 391
Db      392 tatcagaacaacaggggggtcattcaacagctt 425

```

RESULT 9

AAA43182  
ID AAA43182 standard; cDNA: 335 BP.

AC AAA43182;

DT 21-AUG-2000 (first entry)

DE Xenopus secreted expressed sequence tag SEQ ID NO:1922.

KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;  
KW expressed sequence tag; EST; probe; chemokine; proliferative;  
KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;  
KW thrombolytic; antiinflammatory; cytosolic; antibacterial; antifungal;  
KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;  
KW antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;  
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
KW central nervous system disorder; Alzheimer's disease; stroke;  
KW Parkinson's disease; Huntington's disease; coagulation disorder;  
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
KW tumour; infection; depression; psoriasis; ss.

OS Xenopus sp.

PN W0200021990-A1.

PD 20-APR-2000.

PF 15-OCT-1999; 99WO-US24205.

PR 15-OCT-1998; 98US-0104435.

PA (GENY) GENETICS INST INC.





```

KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match          11.1%; Score 163; DB 22; Length 936;
Best Local Similarity 1.3%; Pred. No. 2.2e-27;
Matches 10; Conservative 519; Mismatches 264; Indels 0; Gaps 0;

QY 634 gagagtcacatttcagacatttcagcttggtgataaataagatgac 693
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 795 GGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 736

QY 694 caatgcaagtgataatgaattccaattgacctgagtcagatgtacttc 753
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 676

QY 754 actctgactgactcagtcagtgagatggaatttcagagactgaactgtgaa 813
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 616

QY 814 aaatgaccttccttaactgaagctactttaaatltgaggtctgacaaagaag 873
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 556

QY 874 aggaataatcggttgagtcagacatgaagtgagagtaagtaactcaagaag 933
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 496

QY 934 atgcttcactgaagaagaagcatltaagattttaaaactgtcagaagatccca 993
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 495 WGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 436

QY 994 gaaagttctaatttcataagcaataaagctacatgcaagaataataaca 1053
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 376

QY 1054 gaacactgctctttagatttattgtacttttgctgctggatcggtttaatg 1113
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 316

QY 1114 gacattgtctgacagctcattaaataacaatatattgtcaaaatcgtaactatgc 1173
   : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 315 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 256
QY 1174 ttatttattttaattgtagaagaagaaataatgcctaaataagtttcttcataa 1233
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 255 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 196
QY 1234 atactggaatctgacatggtacaaatlttcttcatactgacagatgatgttaa 1293
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 136
QY 1294 tgaattgagacactgaagttactgaagtgaccttcgaatcaaggatttaagacc 1353
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 76
QY 1354 acaatccttlttaactaactgaatggtctgtttttaaacttgatattccgtatgg 1413
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 16
QY 1414 tgcataattgata 1426
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 15 WWWWWWWWWWWWWWWW 3

```

```

RESULT 13
AAF58259/C
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match          11.1%; Score 163; DB 22; Length 936;
Best Local Similarity 1.3%; Pred. No. 2.2e-27;
Matches 10; Conservative 519; Mismatches 264; Indels 0; Gaps 0;

```



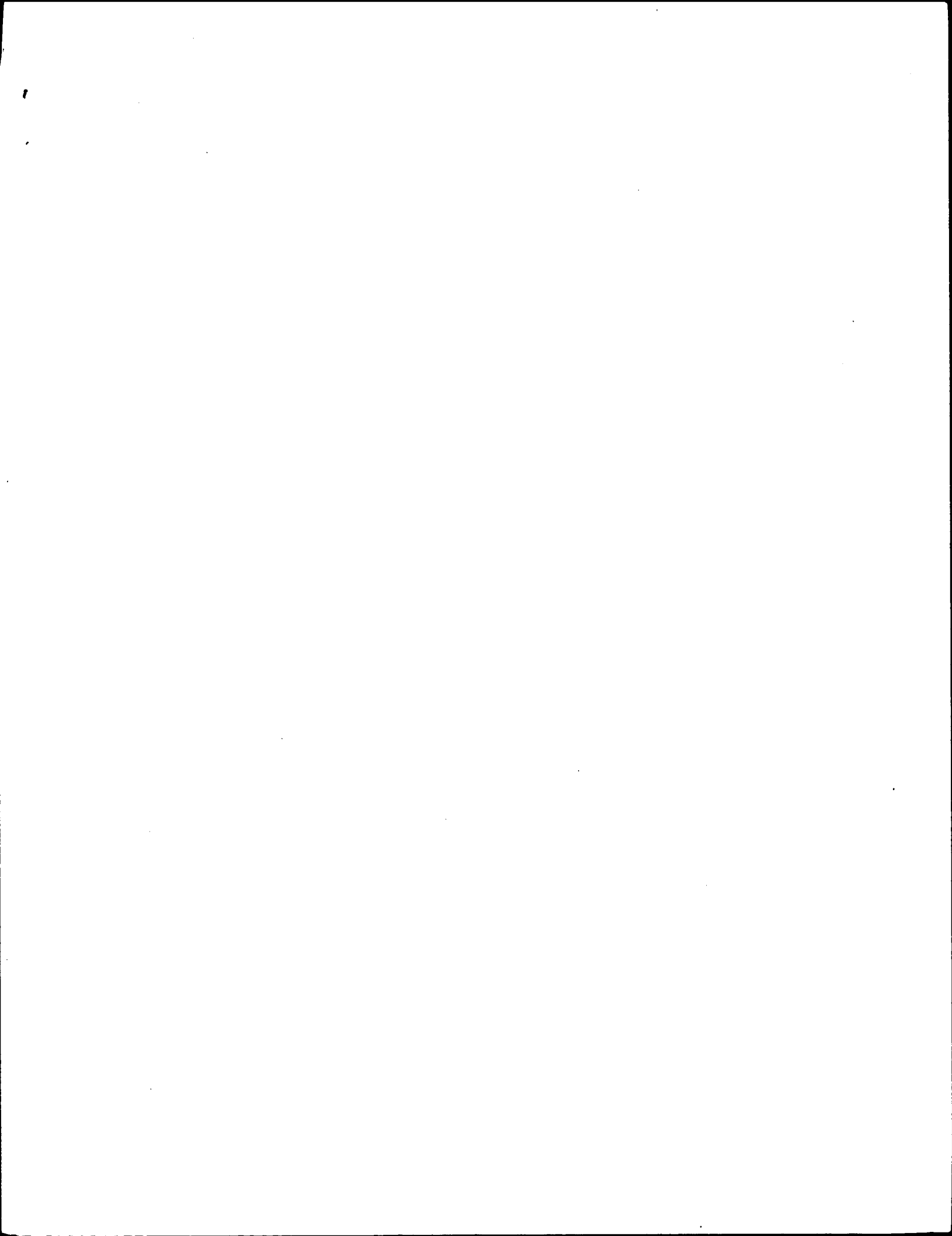


Thu Aug 16 08:48:50 2001

us-09-484-964-1.rng

---

Page 15





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 15:19:49 ; Search time 86.75 seconds  
(without alignments)  
3197.012 Million cell updates/sec

Title: US-09-484-964-1  
Perfect score: 1465  
Sequence: 1 cgggaaggttctgaacc.....gataatgycatgcagcgc 1465

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1286.6	87.8	1514	2	US-09-213-768-1
2	77.6	5.3	473	2	US-08-853-974-2
3	77.6	5.3	473	3	US-09-172-988-2
4	50.6	3.5	7218	1	US-08-232-463-14
5	48.4	3.3	2110	4	US-09-419-459-1
6	46.6	3.2	19124	2	US-08-487-826B-13
7	46	3.1	7218	1	US-08-232-463-14
8	44.8	3.1	87350	3	US-08-781-891-79
9	44.4	3.0	774	3	US-09-014-583-1
10	44.4	3.0	1667	1	US-08-485-284A-1
11	44.2	3.0	1890	6	5312912-3
12	43.4	3.0	1441	4	US-08-821-994-63
13	42.6	2.9	6243	2	US-09-056-075-1
14	42	2.9	1104	4	US-08-307-499-57
15	42	2.9	1104	4	US-09-299-268-57
16	41.8	2.9	2435	4	US-09-306-593-1
17	41	2.8	1582	3	US-08-545-196B-10
18	41	2.8	1582	3	US-08-545-196B-12
19	40.4	2.8	1422	1	US-08-319-704-5
20	40.2	2.7	3933	2	US-08-731-722-3
21	40.2	2.7	3933	2	US-08-731-722-3
22	40	2.7	1198	3	US-09-248-335-27
23	40	2.7	1440	2	US-08-743-637B-174
24	40	2.7	1440	2	US-08-526-840B-174
25	39.8	2.7	2960	3	US-08-913-842-3
26	39.8	2.7	1046	1	US-08-361-467B-4
27	39.8	2.7	1046	1	US-08-484-332C-4

C	28	38.6	2.6	240	1	US-08-628-417-6	Sequence 6, Appl1
C	29	38.6	2.6	1356	4	US-08-675-816-8	Sequence 8, Appl1
C	30	38.6	2.6	2447	2	US-09-014-969-14	Sequence 14, Appl1
C	31	38.6	2.6	2989	6	5378464-1	Patent No. 5378464
C	32	38.6	2.6	4032	1	US-08-107-748-3	Sequence 3, Appl1
C	33	38.6	2.6	4032	1	US-08-245-809-4	Sequence 4, Appl1
C	34	38.6	2.6	4032	5	PCR-US92-01385-3	Sequence 3, Appl1
C	35	38.6	2.6	4358	4	US-08-675-816-7	Sequence 7, Appl1
C	36	38.6	2.6	19124	2	US-08-487-826B-13	Sequence 13, Appl1
C	37	38	2.6	209	1	US-08-146-421-3	Sequence 3, Appl1
C	38	38	2.6	1147	1	US-08-665-716-1	Sequence 1, Appl1
C	39	38	2.6	5718	3	US-08-714-918-48	Sequence 48, Appl1
C	40	38	2.6	5718	4	US-09-265-315-48	Sequence 48, Appl1
C	41	38	2.6	5718	4	US-09-265-315-48	Sequence 48, Appl1
C	42	38	2.6	5718	4	US-09-266-417-48	Sequence 48, Appl1
C	43	37.8	2.6	5852	1	US-07-867-106-2	Sequence 2, Appl1
C	44	37.6	2.6	1183	2	US-08-731-722-8	Sequence 8, Appl1
C	45	37.6	2.6	1332	2	US-09-057-762-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-213-768-1  
Sequence 1, Application US/09213768  
Patent No. 5985664  
GENERAL INFORMATION:  
APPLICANT: Brenda F. Baker  
TITLE OF INVENTION: ANTISENSE MODULATION OF SENTRIN EXPRESSION  
FILE REFERENCE: RTS-0026  
CURRENT APPLICATION NUMBER: US/09/213,768  
CURRENT FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 47  
SEQ ID NO 1  
LENGTH: 1514  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (136)..(441)  
US-09-213-768-1

Query Match	87.8%	Score 1286.6;	DB 2;	Length 1514;
Best Local Similarity	94.1%	Pred. No. 0;		
Matches 1384;	Conservative	0;	Mismatches 74;	Indels 13;
			Gaps	4;
QY	2	gggaagatttgaaccgagaggttctgtcttaaccgagcgctgtgtgga	61	
DB	50	gggaagatttgaaccgagaggttctgtcttaaccgagcgctgtgtgga	109	
QY	62	gaccccggtgaagcgcgtcatctgtctgaccgaggaacacttaactgag	121	
DB	110	gaccccggtgaagcgcgtcatctgtctgaccgaggaacacttaactgag	169	
QY	122	acttgggataaagcaggttgaatatataactcaactcaacttgcagataga	181	
DB	170	acttgggataaagcaggttgaatatataactcaactcaacttgcagataga	229	
QY	182	gtgagattcaactcaagtgaaatgaacacacatctcaagaactcaagaatact	241	
DB	230	gtgagattcaactcaagtgaaatgaacacacatctcaagaactcaagaatact	289	
QY	242	gttaagaagcaggttctcaatgaattcaactcaggttctcttggaggtcagaatg	301	
DB	290	gttaagaagcaggttctcaatgaattcaactcaggttctcttggaggtcagaatg	349	
QY	302	ctgataatcatctccaaaagacttgggaatggaggaagaagtgtgattgattac	361	
DB	350	ctgataatcatctccaaaagacttgggaatggaggaagaagtgtgattgattac	409	



TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 1800 Diagonal Road, Suite 500  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22313-0299  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/232,463  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/935,313  
 FILING DATE:  
 APPLICATION NUMBER: EP 91 114 300.6  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 30472/114 IMM0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703)836-9300  
 TELEFAX: (703)683-4109  
 TELEX: 899149  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: PT9pt-F15

[illegible]

RESULT 5  
US-09-419-459-1  
; Sequence 1, Application US/09419459  
; Patent No. 6222028  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jin-Hao  
; APPLICANT: Cheng, Kuo-Juan  
; APPLICANT: Tsai, Cheng-Fang  
; APPLICANT: Chang, Chia-Chieh  
; TITLE OF INVENTION: CELLULOSE ENZYMES  
; FILE REFERENCE: 08919/037001  
; CURRENT APPLICATION NUMBER: US/09/419,459  
; CURRENT FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2110  
; TYPE: DNA  
; ORGANISM: Pityomyces rhizinflata  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)...(1749)  
US-09-419-459-1

Query Match 3.3%; Score 48.4; DB 4; Length 2110;  
Best Local Similarity 47.4%; Pred. No. 0.0073;  
Matches 145; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 950 aaagcctttaaagattttttaaataatcttgcaagaatccagaagaagttctaat 1009  
DB 1720 aatggtatggtatggtcgcgataatggttaataataagaataaataaattt 1779  
QY 1010 catgaagcaataaagcgcatacgagaaatgataaacaagaacactgctcttt 1069  
DB 1780 aatgaataatattttaaataaataaataaataaataaataaataaataaata 1839  
QY 1070 agatttatttgaatttggcctggagataatggttttaagagacatgctgtacca 1129  
DB 1840 aatgtcaattaaattttatttcttatttattttaaataaataaataaataaaga 1899  
QY 1130 gcttcattaaataaataatctgcacaaatcgtaactaagcttatttatttatt 1189  
DB 1900 aatataaataaataaataaataaataaataaataaataaataaataaataaata 1959  
QY 1190 gataagaagaaaaaaatgcctaaataaaggtttcttgatcaataactggaatggc 1249  
DB 1960 aagcaaaaaaaagatttaataaataaataaataaataaataaataaataaata 2019  
QY 1250 atggtta 1255  
DB 2020 ataata 2025

RESULT 6  
US-08-487-826B-13/C  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhuan  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 3.2%; Score 46.6; DB 2; Length 19124;  
Best Local Similarity 49.4%; Pred. No. 0.049;  
Matches 121; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 378 tcatcaacagtttgatattcttttttttttttttttttttttttttttttttttt 437  
DB 15856 TTAATTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 15797  
QY 438 tttaaaagttcttcttgtaattggtgtcacaagaacgaaatgaagaacgaccccat 497  
DB 15796 TATGTATATATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 15737  
QY 498 ctcttggaacatcgggaattgaattcagtgctcaattatcattatggtgtttc 557  
DB 15736 TTAATTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 15677  
QY 558 atgtgctgattttgtgtgacaaagcctcagccctcatattaccctctcttttaa 617  
DB 15676 AATGTTTTTTTTCTCTCTTTGTTTAAATTTTAAATTTTAAATTTTAAATTTT 15617  
QY 618 aatt 622  
DB 15616 AAATT 15612

RESULT 7  
US-08-232-463-14/C  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA

Thu Aug 16 08:48:50 2001

us-09-484-964-1.rn1

Page 5

ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgt-Fls  
US-08-232-463-14

Query Match 3.1%; Score 46; DB 1; Length 7218;  
Best Local Similarity 2.7%; Pred. No. 0.046; Indels 0; Gaps 0;  
Matches 10; Conservative 212; Mismatches 152;

4 gaagattgttaaccccgagcaggtctgctaccgagccgctgctgagaga 63  
1448 GAAGATTGTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1389  
64 ccccggtgagcagcagcagcagcagcagcagcagcagcagcagcagc 123  
1388 RRR 1329  
124 ttggggataagagagagagagagagagagagagagagagagagagag 183  
1328 RRR 1269  
184 gagatcactcaagtgaaatgacaacatctcaagaagaactcaaacatc 243  
1268 RRR 1209  
244 caag 303  
1208 RRR 1149  
304 gataatcactcaagagagagagagagagagagagagagagagagagag 363  
1148 RRR 1089  
364 gaacaacgagagag 377  
1088 RRR 1075

RESULT 8  
US-08-781-891-79/C  
Sequence 79, Application US/08781891  
Patent No. 6090620  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui

APPLICANT: Yu, Chang-En  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
TITLE OF INVENTION: WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-79

Query Match 3.1%; Score 44.8; DB 3; Length 87350;  
Best Local Similarity 5.1%; Pred. No. 0.25;  
Matches 108; Conservative 0; Mismatches 87; Indels 1; Gaps 1;

182 gtgagatcactcaagtgaaatgacaacatctcaagaagaactcaaacatc 241  
12134 GTGAGATTGAAGTGAACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 12075  
242 gtcaag 301  
12074 GTAAGGGGACGGTTTGCCATGAGGAGATGAGTTCAATTGACAGGAGCC 12015  
302 ctgataatcactcaagagagagagagagagagagagagagagagagag 361  
12014 ATGAACAGACACATCTGCACATTGGAAGT-GAGCCTTAAGCAAAAGATGT 11956  
362 aggaacagagagag 377  
11955 AGCAGCAGACATGGAG 11940

RESULT 9  
US-09-014-583-1  
Sequence 1, Application US/09014583  
Patent No. 6140077  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, Takeshi  
APPLICANT: SUZUKI, Tadashi  
APPLICANT: TOKUDA, Junko  
APPLICANT: KATO, No. 6140077uo  
APPLICANT: SAKAI, Yasuyoshi  
APPLICANT: MOCHIZUKI, Daisuke  
APPLICANT: TAKAHASHI, Hitoshi  
TITLE OF INVENTION: METHOD FOR PRODUCING PHYTASE

GENERAL INFORMATION:  
APPLICANT: SAKAI, YASUYOSHI  
APPLICANT: TANI, YOSHIKI  
APPLICANT: SHIBANO, YUJI  
APPLICANT: KONDO, HIROTO  
APPLICANT: HATANAKA, HARUYO  
TITLE OF INVENTION: NOVEL VECTOR HAVING PROMOTER THAT IS  
TITLE OF INVENTION: INDUCIBLE BY METHANOL AND/OR GLYCEROL  
NUMBER OF SEQUENCES: 18

RESULT 10  
US-08-485-284A-1  
; Sequence 1, Application US/08485284A  
; Patent No. 5750372

[illegible]

Query Match	Score	DB #	Length
3.0%	43.4	4	1771

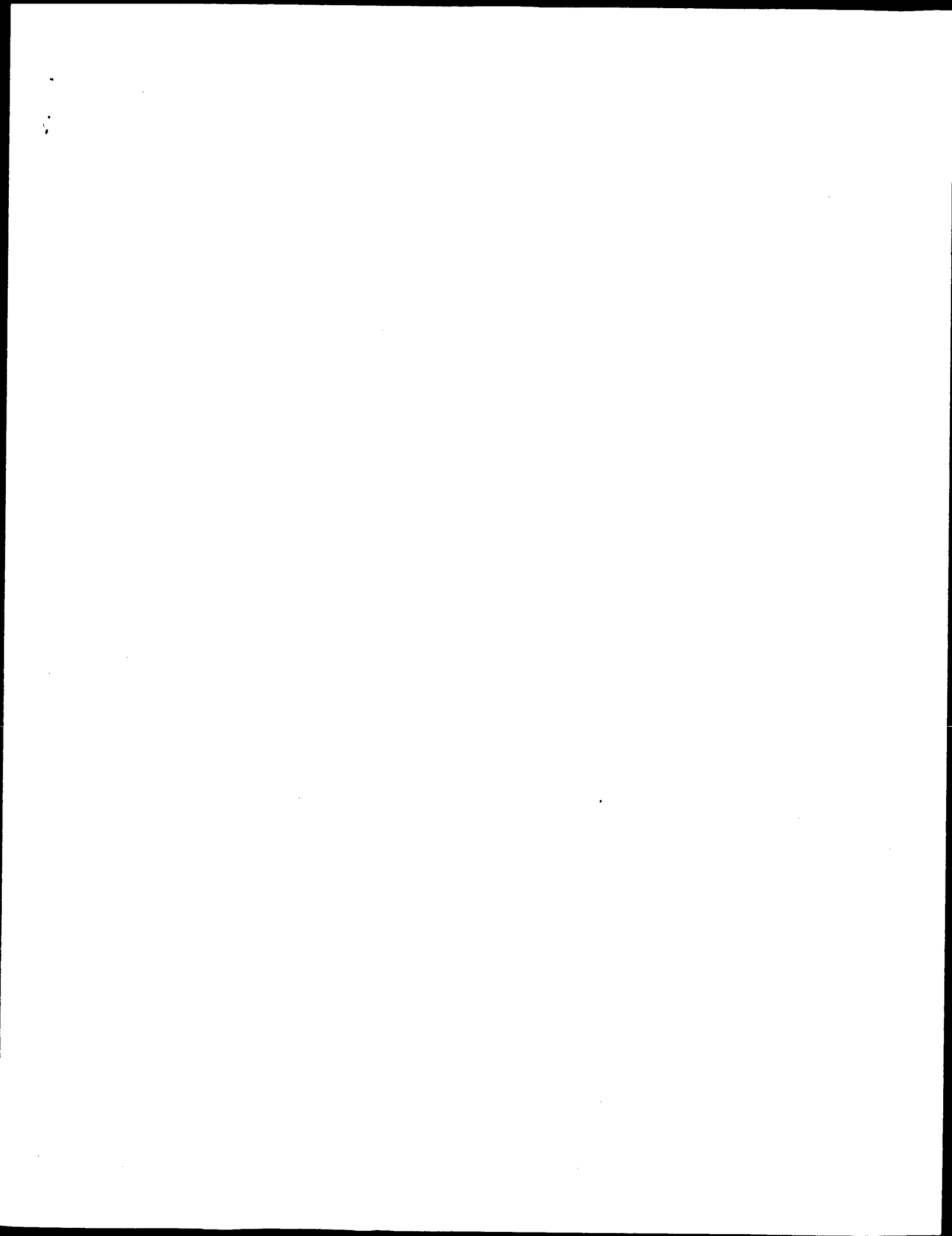
Db 1408 AATAATTTTATATTTTTTAATTTTTTAATTTTATATTATATA 1349

Query Match	2.98;	Score 42;	DB 1;	Length 1104;
Best Local Similarity	47.78;	Pred. No. 0.21;		

INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 15:18:29 ; Search time 1167.5 Seconds  
(without alignments)  
11861.611 Million cell updates/sec

Title: US-09-484-964-1  
Perfect score: 1465  
Sequence: 1 cgggaagatttgtaaaccc.....gataaatggcattgcacgcc 1465

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	945.2	64.5	1046	AL528799	AL528799 AL528799
2	933.2	63.7	998	AL563251	AL563251 AL563251
3	864.2	59.0	924	AL565221	AL565221 AL565221
4	852.8	58.2	106	AL521484	AL521484 AL521484
5	846.8	57.8	105	AL533666	AL533666 AL533666
6	843.8	57.5	106	AL583458	AL583458 AL583458
7	842.4	57.5	106	AL570877	AL570877 AL570877
8	840.6	57.4	106	AL549442	AL549442 AL549442
9	800.4	55.6	105	AL521483	AL521483 AL521483
10	779.6	53.2	933	AL583408	AL583408 AL583408
11	760.8	51.9	1175	AK002536	AK002536 Mus muscu
12	751.2	51.3	1180	AK011074	AK011074 Mus muscu
13	737	50.3	943	AL532393	AL532393 AL532393
14	734	50.1	1166	BG169405	BG169405 BG169405
15	732	50.0	804	BG540101	BG540101 BG540101
16	730	49.8	835	BG541610	BG541610 BG541610
17	724	49.4	822	BG527341	BG527341 BG527341
18	716	48.9	846	BG529400	BG529400 BG529400
19	714.4	48.8	756	AL532394	AL532394 AL532394
20	714.4	48.8	827	BG531082	BG531082 BG531082
21	714.2	48.6	798	AL538719	AL538719 AL538719
22	711.6	48.6	731	AL5066583	AL5066583 ova47f05.x
23	702.2	47.9	979	BE738302	BE738302 BE738302
24	694.8	47.4	762	AL037282	AL037282 DKEP564A
25	693.8	47.4	1045	BE887652	BE887652 BE887652
26	691.2	47.2	812	BG505745	BG505745 BG505745
27	690.2	47.1	786	BG429188	BG429188 BG429188
28	685	46.8	816	BG533601	BG533601 BG533601
29	684.2	46.7	858	BE888192	BE888192 BE888192
30	684	46.7	763	BG535057	BG535057 BG535057
31	683	46.6	781	BG502205	BG502205 BG502205
32	682.2	46.6	752	BG531369	BG531369 BG531369
33	678.6	46.3	761	BG547743	BG547743 BG547743
34	673.8	46.0	797	BG503010	BG503010 BG503010
35	672.4	45.9	858	BG530869	BG530869 BG530869
36	665.8	45.4	915	BG391730	BG391730 BG391730
37	664.8	45.4	733	BG500627	BG500627 BG500627
38	663	45.3	1148	C88594	C88594 C88594
39	649	44.3	760	AM006623	AM006623 wt06a11.x
40	645	44.0	769	BG528729	BG528729 BG528729
41	641.2	43.8	716	AV714091	AV714091 AV714091
42	639.6	43.7	716	BG502885	BG502885 BG502885
43	633	43.2	791	BG248994	BG248994 BG248994
44	627.8	42.9	770	AU127846	AU127846 AU127846
45	624	42.6	801	BG532372	BG532372 BG532372

## ALIGNMENTS

RESULT	1	AL528799	1046 bp	mRNA	EST	13-FEB-2001
LOCUS	AL528799	LTI_NFL001_NBC4	Homo sapiens	CDNA clone	CSDD001YH20	5
DEFINITION	AL528799	Prime, mRNA sequence.				
ACCESSION	AL528799.1	GI:12792292				
VERSION	EST.					
KEYWORDS	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 1046)					
AUTHORS	Ll.W.B., Gruber,C., Jessee,J. and Polayes,D.					
TITLE	Full-length cDNA libraries and normalization					
JOURNAL	Unpublished (2001)					
COMMENT	Contact: Genoscope					

Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

## source

1. 1046  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDD001YH20"  
/clone\_11b="LTI\_NFL001\_NBC4"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestechnology.com URL : http://fulllength.invitrogen.com"

BASE COUNT 310 a 185 c 231 g 318 t 2 others

## ORIGIN

Query Match 64.5%; Score 945.2; DB 106; Length 1046;  
Best Local Similarity 98.7%; Pred. No. 2e-189;  
Matches 983; Conservative 1; Mismatches 9; Indels 3; Gaps 3;

QY	2	ggaagagattglaaaccccgagagaggttcgttaccgagcgctgtcgga	61
DB	33	GAGAAGATTGTGTAACCCCGAGGAGGCTTCTGTTACCGAGCGCTGTGCGGA	92
QY	62	gaccccgagtgaaagccatcatcatgtctgacagagagagaaacactcaatg	121
DB	93	GACCCCGGATGAAAGGAGGATGATTAATTAATCAAGTCTTGAAGGAGGAGG	152
QY	122	acttgaggatagaagagaggttgatataataactaaagtcattgagagataca	181
DB	153	ACTTGGGAGTAAAGGAGGATGATTAATTAATCAAGTCTTGAAGGAGGAGG	212
QY	182	gtgagatcactcaagtgaaatgacaacacatcacaagaactcaagaatactat	241
DB	213	GTTAGATTCTCAAGGAGGAGGATGATTAATTAATCAAGTCTTGAAGGAGG	272
QY	242	gtcaagagacaggtgtccatgaatcaactgaagttctcttgaaggtgaagaatg	301
DB	273	GTCAGGAGGAGGAGGAGGATGATTAATTAATCAAGTCTTGAAGGAGGAGG	332
QY	302	ctgatatcatatctccaaagacagggagatgaggaagaagatgtgattgaattac	361
DB	333	CTGATATCATCTCCAAAGAGGAGGAGGATGATTAATTAATCAAGTCTTGA	392
QY	362	aggaacaaacgggggtcattcaacagtttagatatcttttattttttttttccc	421
DB	393	AGGAACAAACGGGGGTCTATCAACAGTTTGAATTTCTTTTATTTTCTTTCC	452
QY	422	tcaatcctt	481
DB	453	TCAATCTTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	512
QY	482	aaacatgaccccatctcttgaaacacatcgttaattgaaattcagtgatcattac	541
DB	513	AAACATGACCCCATCTCTTTGAACATCTGTAATTTGAATTTGATGATTTATTC	572
QY	542	attatgtt	601
DB	573	ATTATGTTTGTGTTTATGTTGTTGTTTGTGTTGTTGTTGTTGTTGTTGTTG	632
QY	602	accttccttt	661
DB	633	ACCTTCCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTG	692

OY	662	tttccaggtctggtggatataaatgaagtgcacaaatcgaatcgatgfttcataatgactttcca	721	
Db	693	TTTTTCAGGCTTGCGTGAATTAATAAGATGCACCAATCAAAAGTGTTCWAAATGACTTTTCCA	752	
OY	722	attgcccttgatgltc-agcatlgtatctaactcaactcctctgactgactgtaacttcagttgga	780	
Db	753	ATTGGCCCTGATGTTCAGCATGTGATTACTTCACCTCCTGAGCTGATGACTTTTCAGTGGGA	812	
OY	781	gatgagaattttcacagaacctgaactcgttggaaaaatagacccttctctaacttcgaagcta	840	
Db	813	GATGGAAGTTTTCAGAGAACCAGACGTGGAAAAAAGACCTTTCCCTTAACCTGMAACTGA	872	
OY	841	ctttaaaattttagagggtcgcgaccacaagaagaagatatcaatgattgtaagtaagaatga	900	
Db	873	CTTTTAAAATTTCAGGGGTCTGAGCACCAAANAAGAGATATTCAGTTGAAGTCAAGATGA	932	
OY	901	cagataagtgtagaagtaatgaactaacccaagaatgagcttcaactgaagaaaagcatlitt	960	
Db	933	CAGATTAAGGTGAAGAGTATGACACTAACCTCCAAAGATGGC-TYACTGAAGAAAAGCATTTT	991	
OY	961	aagatttttaaataatcgttcagaagatcccagaa	996	
Db	992		1026	
RESULT	2			
LOCUS	AL563251/C	998 bp	mRNA	
DEFINITION	AL563251 LTL_NFL001_NB.C4 Homo sapiens cDNA clone CSDD0001YH20 3	EST	16-FEB-2001	
ACCESSION	AL563251	prime mRNA sequence.		
VERSION	AL563251.1	GI:12912463		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 998) Li,W.B., Gruber,C., Jessee,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr, Location/Qualifiers 1..998			
FEATURES	Source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="CSDD0001YH20" /clone_lib="LTL_NFL001_NB.C4" /sex="male" /tissue_type="neuroblastoma cells" /lab_host="DH10B" /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-Oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filiangeli@tech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	336 a	193 c	160 g	
ORIGIN		298 t	11 others	

QY	152	tttaactcaaaagcattctggacagaggtatgaagtgaaattcacttcaaaatgtaaaatgcaa	211
Db	998	TTAACTCAAAAGCATTTGACAGATGGCAGTGCAGATTCCTCTCAAAAGTARATGACAA	9339
QY	212	caaatctcaagaagaaactcaagaatcatcatctgccaagaagagtggttccaatgtaatcac	2711
Db	938	CACTTCTCAAGAAACCTCAAAAGAGTCGAAATGTCAAAGACAGGGGTGTTCCAAATGATTTAC	879
QY	272	tcaagttctctctttagggctcagaagaaattgctgcataatcatcaccccaaaagaaactggaa	3331
Db	878	TCAGTTCCTCTTTGGGGTGCAGAAATTTGCTRATATCATATCTCCAAAGRACOTGGGAA	8119
QY	332	tggaggaagaagaatgtagttgaagtttacaagaaacaaacggggggtgcatcaacagttc	331
Db	818	TGGAGGAACAAATGATGATGAAAGTTTTCAGAGAACAAACGGGGGGTCATTCACACATTTT	759
QY	392	agatactcttcttaattttttctttccccaatcccttttttttttttttttttttttttttttt	451
Db	758	AGATATCTCTTTTATTTTCTTTTCTTTTCCCTTAATCCCTTTTATTTTAAAAATAGTTC	659
QY	452	ttttgttaatgtagtggttccaacaaacggaattgaacacatgagccacatctcttgaacatc	511
Db	698	TTTTGTANTGTGTCCTTCAAAACGGAATTTGAAGATCGACACCCATCTCTTTGAAACNTC	639
QY	512	tggtaattggaattcagtgctcatatcatatatttttgttttaatttgctgattt	571
Db	638	TGGTATATTGAATTTCTAGTCTCATTTATTCATATATGTGTGTTTTCATTTGCTGATTTT	579
QY	572	tgggtacaagcctcaagtcctccctcaataataccctccctttttaaanaatlaagtgtaga	631
Db	578	TGTCCTCAAGCCTCAAGTCCCTTCATATTTACCTCTCTTTTAAAAATTTAGCTGNGA	519
QY	632	cagagaaggtcaaccttttcaagagacattgcaatttcaagcttggtggtatataaagaatg	691
Db	518	CAGAGGGGTCACTTTTCAGAGAAATTGCAATTTTCAGGCTTGtGGATTAATAAGATCG	459
QY	692	accaatgcgaagtgtcatatagacttccaaattgagccctgagtctc- agcatgtagtac	750
Db	458	ACCAATGCAAGTGTTCATAAAGATCTTCCATTTGGCCCTGATGTTCGATCATGTGATTC	399
QY	751	ttcatctctgacagtgacttcaagtgaggagatggaagtttccagaagactgaactgtg	810
Db	398	TTTACTCTCGACGTCTACTTTCAGTGGAGATGGAAGTTTTCAGAGGACGTCAACTGTG	339
QY	811	gaanaatgaccttcccttaacttgaagctacatctttaaanaatttgaaggtctgagccaaag	870
Db	338	GAAAAATACCTTTCTCTTACTTGAACCTCTCTTTAAAAATTTGAGGGTCTGAGCCAAAG	279
QY	871	aagaagaatcatcaggttgaagaatgacaagatgacagatagtggaagatgaactaactcca	930
Db	278	GAGAGGAATATTCAGGTTGGAGTCAAGATTAACAGATTAAGTGAAGATTAATTAATCTCA	219
QY	931	aagatggtctactgaagaaagagcatctttaaagattttaaanaactctgtcagaagatc	990
Db	218	AAGATGCTTTACTGAGAAAGGATTTTAAATTTTAAAAATCTTGTCAGAAATC	159
QY	991	ccaaaaaagttctaatttctattagcaataataaagctatatagcagaagaatgatata	1051
Db	158	CGAGAAAGTCTAATTTTATTTGCAATTAATAAAGCTATATCATGCACAAAATGATACOA	99
QY	1051	acagaacacgtctcttctttagatttatgttgaacttttggcctggagatatggttttaa	1110
Db	98	ACAGAAACAGTCTCTTTT- GATTTTATTGTACTTTTGGCCTGSMATATGGGTTTTTA	40
QY	1111	atggacattgtctgtacagctctcaataataaacaat	1149
Db	39	ATGACATTTGCTGTACACCTTCATTTAAAAATAACAAT	1

Query Match	63.7%	Score 933.2	DB 106	Length 998
Best Local Similarity	96.5%	Pred. No. 6.9e-187		
Matches 964	Conservative	8	Mismatches 25	Indels 2
				Gaps 2

RESULT	3				
LOCUS	AL565221/c	924 bp	mRNA	EST	16-FEB-2001
DEFINITION	AL565221 L1-F1013-Fbrn1 Homo sapiens cDNA clone CSDF0020Y03 3				

ACCESSION	prime, mRNA sequence.
VERSION	AL565221
KEYWORDS	AL565221.1    GI:12916380
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	1 (bases 1 to 924)
COMMENT	LI,W.B., Gruber,C., Jesse,J. and Polyes,D. Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES	Location/Qualifiers

BASE COUNT	326 a	163 c	142 g	288 t	5 others
ORIGIN					

Query Match	59.08;	Score	864.2;	DB	106;	Length	924;
Best Local Similarity	98.28;	Pred	No. 2.5e-172;				
Matches 910; Conservative	5;	Mismatches	8;	Indels	4;	Gaps	4

[illegible]





```

Db      464 TCAATCCTTTTATTTTAAATAAGTCTTTTATGTTGTTGCAAAAGGAATTG 523
QY      482 aaaaacggacccaccatctcttgaacatcgtgtaattgaattcgaatcattatc 541
Db      524 AAAACGACACCCCATCTCTTTGAAACATCTGTATTTGAATTTCAAGTCATATTTC 583
QY      542 attatgttcttcttcaatctgctgatttttgatgaacagcccaagccctcaatc 601
Db      584 ATATATGTTGTTTCAATGTCGTGATTTTGTGATCAAGCCCTGATCCCTCATATT 643
QY      602 accctctcttctttaaataatcgtgtgacagagagcgcaccccttttgaagacatg 661
Db      644 ACCCTCTCTTTTAAATAATAGTGTGACAGAGAGTCCCTTTTCAAGACATGCA 703
QY      662 ttctcagccttggtgtgataaataagatcgaacatcgaatgctcataatcattcca 721
Db      704 TTTTCAGGCTTGCTGTGATTAATTAAGATGACCAATGCAAGTTTCATATGACTTTCCA 763
QY      722 attgcccctgattgtc-aggatgtgattactcactcctgagctgtgacttcaatgga 780
Db      764 ATGGCCCTGATGTCTACATGTGATTAATTAATTAATTAATTAATTAATTAATTA 823
QY      781 gatgagaagtttttcaagagacgaactgtggaataatgaccccttcttgaagcga 840
Db      824 GATGGAAGTTTTCAGAGAACTGAATCTGGAATAATGAACCTTCTTAATCTGAAGCTA 883
QY      841 ctttaaaattgaaggtctgac 864
Db      884 CTTTAAATTTAGAGGCGCTGRC 907

```

```

RESULT 6
LOCUS   AL583458 913 bp mRNA EST 16-FEB-2001
DEFINITION AL583458 L1.NFL010.BC2 Homo sapiens cDNA clone CS0D1012YD05 5
ACCESION AL583458
VERSION AL583458.1 GI:12952440
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 913)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.
          Location/Qualifiers
            1. 913
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="CS0D1012YD05"
              /clone_1b="L1.NFL010.BC2"
              /sex="male"
              /tissue-type="B cells from Burkitt lymphoma"
              /note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA
              was primed with a NotI-oligo(dT) primer. Five prime end
              enriched, double-stranded cDNA was digested with NotI and
              cloned into the NotI and EcoRV sites of the pCMVSPORT 6
              vector. Library was normalized. Library was constructed by
              Life Technologies. Contact: Feng Liang Life Technologies,
              a division of Invitrogen 9800 Medical Center Drive
              Rockville, Maryland 20850, USA Fax: (1) 301 610 8371
              Email: filiang@lifestech.com URL:
              http://fulllength.invitrogen.com"

```

```

BASE COUNT 268 a 166 c 199 g 279 t 1 others
ORIGIN

```

```

Query Match 57.6% Score 843.8 DB 106 Length 913;
Best Local Similarity 98.9% Pred. No. 5e-168 7; Indels 3; Gaps 2;
Matches 871; Conservative 0; Mismatches
QY 15 aaaccggagcggaggtctgcttaccgaagccgcgtgctgctgagaccccggtga 74
Db 1 AAACCCGGAGCGAGGTTCTCTCTACCCGAGCGCGTGTGGAGACCCCGGGTGA 60
QY 75 agccacgctcatcgtctgaccagaggaacaaacctcaactgaggaacttgaggataa 134
Db 61 AGCCACGCTCATCATGTCGTACACAGAGGCAAAACCTTCAACTGAGGACTTGGGGATAA 120
QY 135 gaagagaagtgataataataaactcaagatgagcagagatagcagatgactcact 194
Db 121 GAAGGAAGGTGAATATTAATAAATCAAGTCAAGTATGAGATGACATGATTCATT 180
QY 195 caaagtgaatgaacaacacatcacaagaactcaagaatcactgctcaagaacagag 254
Db 181 CAAAGTGAATGACACACATCTCAAGAACTCAAGAAATCATATCTGTAAGACAGAGG 240
QY 255 tgttccaatgaatcactcaggttctcttgaagggtcagaagaattgctgatactac 314
Db 241 TGTTCAATGAATTCATCAGGTTCTCTTGAAGGCTCAGAAATGCTGATTAATCATAC 300
QY 315 tccaaagaactgggaatggaggaagaagatgtaattgaatttcaagacaacaggg 374
Db 301 TCCAAAGAAGTGGGAATGAGAGGAAGAAAGATGATTAATTAATTAATTAATTAATTA 360
QY 375 ggtcattcaacagtttagatattcttcttcttcttcttcttcttcttcttcttctt 434
Db 361 GGGTCATTCACAGTTTATGATATCTTTTATTTTATTTTCTTTTCCCTCATCTTTT 420
QY 435 attttaaataagttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 494
Db 421 ATTTTAAATAATGATCTTTTGTGTAATGAGTGTCAAAACGGAATGAAATGAGCAACC 480
QY 495 catcctcttgaacaacatcgtgaattggaattggaattggaattggaattggaattg 554
Db 481 CATCTCTTGAACATCTGTGTAATGATTAATTAATTAATTAATTAATTAATTAATTA 540
QY 555 ttcattgtgctgatttcttcttcttcttcttcttcttcttcttcttcttcttcttct 614
Db 541 TTCAATGTCCTGATTTTGTGATCAAGCTCAGTCCCTCATATTAATTAATTAATTAAT 600
QY 615 taaaatacgtgtgacagaagaagttcaacttctcaggaacttgacttctcaggtgtg 674
Db 601 TAAAAATTAAGTGTGACAGAGAGGTCACCTTTTCAAGACATTCATTTTCAGGCTGT 660
QY 675 gttgataaataagatcgaacaaatgcaagtgttcataatgacttccaatgtccctgag 734
Db 661 GGTGAATAAATAAGATGACCAATGCAAGTGTTCATTAATGACTTTCCATTTGGCCGTG 720
QY 735 ttc-aggatgtattacttactcactcctgagctgacttgaacttcaagtggagatgtt 793
Db 721 TTCTPACATGATTAATCTTCACTCCCTGAGCTGTGACTTCACTGAGAGTGAATTTT 780
QY 794 cagagaactggaactgtgaaataatgaccccttcttcaacttgaagttactttaaattg 853
Db 781 CACAGAACGGAATGTGGAATAATGACCTTTCTTAATCTGAAGACTTTTAAATTTG 840
QY 854 aggtctgagcaaaagaagaagaatcaggttgagca 894
Db 841 AGGCTCTGCAC -AAAAGAAGAGATATCAGGTGAAGTCA 879

```

```

RESULT 7
LOCUS   AL570877 887 bp mRNA EST 16-FEB-2001
DEFINITION AL570877 L1.NFL006.PL2 Homo sapiens cDNA clone CS0D1012YA10 3
ACCESION AL570877
VERSION AL570877.1 GI:12927614

```

KEYWORDS	EST.
SOURCE	
ORGANISM	
human.	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
1 (bases 1 to 887)	
Li, W. B., Gruber, C., Jesse, J. and Polyes, D.	
Full-length cDNA libraries and normalization	
Unpublished (2001)	
Contact: Genoscope	

FEATURES	Location/Qualifiers
source	1. .887

BASE COUNT	270 a	177 c	149 g	274 t	17 others
ORIGIN					

Query Match	57.5%;	Score 842.4;	DB 106;	Length 887;
Best Local Similarity	97.3%;	Pred. No. 9	8e-168;	
Matrices	863: Conservative	13: W		

OY	111	ttcaactgaaggaacttgagggaataagaaacgaagtgaaatataatgaactaaagcatg	170
Db	887	TCCAACTGAGGACTTGGGGGTAAGAAAGAGGCGTAATATATWAGACTCAAGCATTTGG	828
OY	171	a-caagataagcagtgagaattcaacttcaaaatgaaatgacaacacalcacaagaactca	229
Db	827	ACCAAGGAATGAGAGGAGTTCCTCTCAAGGAAATGACAAACACATCTCAAGAACTCA	768
OY	230	aagaatactctgcaagacagagtgattcaatgaattacatcagttctcttgag	289
Db	767	AAGAATCATCTGTCCAAAGACAGGGGTCTCCATGCAATTCACTGAKTTCTCTTTAGG	708
OY	290	gtcagaagaattgctgataatcatatcccaaaagaactcgggaatgaggaagaaatgta	349
Db	707	GTCAGAGAAATTCGAAATATCATCTCCAAAGACAGCGGAATGAGAGGAAGATGTGA	648
OY	350	ttgaagttatcaagaaacaacggaggagtcattcaacaagtttgatactctttttttt	409
Db	647	TTGAAGTTATACAGAACAAACGGGGGTCAATTCACAGATTGAAATTTCTTTTATTTT	568
OY	410	ttctctttccctcaatccctttttatattttaaataagttctcttgytaatgfgtgtc	469
Db	587	TTTTCTTTTCCCTCAATCCCTTTTATTTTAAAAAATGATCTTTGTATGSGGTTC	528
OY	470	aaaacggaaatgaaactcggaccccaatcctttgaaacatcggtaatttgaaatctag	529
Db	527	AAAACGGAATTGAAAACCTGGCACCCCATCTCTTTGAACACATCGGTATTTGAATCTAG	468
OY	530	tgcataatcatcattatgttgtttcaattgctgcataattctgfgatcaaacctcagt	589
Db	467	TGCTCATTTATTCATTATTGTGTGTTTCATTGTGTCGATTTTTGGTGATCAACCTCAG	408
OY	590	ccccctaatatcccccctctttttaaaataacgctgacagaagaggtcaacctttt	649
Db	407	CCCCCTATATACCCCTCCCMCTCHVAAAAATTAAGTGTGCACAGAGAGGTCAACCTTTT	348

OY	650	caggacattgcatcttcaaggctcttgatgaataaataagatgccaaatgcgaatggttc	709
Db	347	CAGGACATTTGCATTTTCCASGCTTGGTGATAAATAAGATGCACCAATGCAGAGTTCAT	2888
OY	710	aatgaccttccaattggccctgaatctc-agcagctgattacttcaactcctgagctgtga	7688
Db	287	AATRACHTMCCAAATWGGCCGTGATGGTTCTACACATGTGATTACTTCAACYCCGTGACTGTTRA	2278
OY	769	cttccagtgggagatgtagaagctttctcagagaactgaactcgttgaaaaatgaccttcct	8278
Db	227	CTRTCAGTGGGAGATGGAGAGTCTHTCAGAGACAGCACTGTGGAAAAATGACCTTCTCT	168
OY	829	aacttgaagctactctttaaataattgtagygtcttgcaccaaagaagaggaataatcaggttg	8888
Db	167	AACCTTGAAGCTACTTTTTAAAAATTTTGAAGGCTGTGCACCCAAAAAGAAAGAAATATCAGGTTG	108
OY	889	aagtcacaagatgacagataaagtgtagtaatgactaactctcaaaagtatggcttaactgaag	9478
Db	107	AAGTCACAGATACGATRAAGGTGAGAGTAATGACTTDCCTCAAAAGATGGCTTCACTGAAG	48
OY	949	aaaagcgattttaagattttttaaaaaactcgttcaagaagatccaga	995
Db	47	AAAAGGCATTTTAAAGATTTTTTAAAAATCTGTGCAGAAAGATCCCAN	1

[illegible]

```

REFERENCE
  1. (Phases 1 to 898)
AUTHORS
  I. W. B., Gruber, C., Jesse, J. and Polyes, D.
TITLE
  Full-length cDNA libraries and normalization
JOURNAL
  Unpublished (2001)
COMMENT
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
FEATURES
  Location/Qualifiers
  1..898
source

```

BASE COUNT	259 a	160 c	200 g	275 t	4 others
ORIGIN					

Query Match	57.4%;	Score 840.6;	DB 106;	Length 898;
Best Local Similarity	98.3%;	Pred. No. 2.4e-167;		
Matches 876;	Conservative 4;	Mismatches 8;	Indels 3;	Gaps 3;

QY 2 gggaagagatttgttaaaccccgagagcgaggtctctgtctaacccgagagccgctgctgtgcgga 61

```
Db 9 GAGAAAGATTTTAAACCCCGAGGAGTTCCTGCTAACCCGAGGCCCTGCTGCGGA 68
Qy 62 gaccccgaggtagaagccacgcgtcalcatgctgacaggaagcaaaccttcaactgagg 121
Db 69 GACCCCGGGGTGAACCCACCCGATCATGCTGACACGAGGAGCAAAACCTTCAACTGAGG 128
Qy 122 acttgggggataagaagcaagtgatataataactaaagtcattggaaggttaaca 181
Db 129 ACTTGGGGGATTAAGAGGAAGGTGAATATATTAATTAATTAATTAATTAATTAATTA 188
Qy 182 gtgagattacttcaaatgaaatgaaacacacacacacacacacacacacacacacacac 241
Db 189 GTGAGATTCTTCAAGAGAAATGACACACACTCTCAAGAACTCAAGAAATCATACT 248
Qy 242 gtcaaaagcaagggtgttccaatgaaatcaactcaaggttcttcttgaaggtcagagaat 301
Db 249 GTCAAGAGAGAGGCTGTCATGATGATTCACCTGAGTTCTCTTGAAGGTGAGAGAAAT 308
Qy 302 ctgataatcatatctccaaagaacatgaggaatgaggaagaagaatgattgattatc 361
Db 309 CTGATATATATCTCCAAAGAACTGGGAATGAGAGAAAGATGATGATGATGATGATGAT 368
Qy 362 aggaacaacaggggggttcaacacagttagatacttcttcttcttcttcttcttctcc 421
Db 369 AGGACAAACGAGGAGGTCTTCAACAGTTAGATATCTTTTATTTTCTTTTCTTTTCC 428
Qy 422 tcaatccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctcc 481
Db 429 TCAATCTCTTTTATTTTAAATAGTTCTTTTGAATGATGATGATGATGATGATGATGAT 488
Qy 482 aaaaacggacacccatcttcttgaacacatgtaattgtaattgtaattgtaattgtaatt 541
Db 489 AAAACGGACACCCATCTTTGAACATCTGTATTTGAATTTGATTTGATTTGATTTGATTT 548
Qy 542 atattgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 601
Db 549 ATTATGTTGTTTTCATGTTGCTGATTTTGTGATTTTGTGATTTTGTGATTTTGTGATTT 608
Qy 602 acccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 661
Db 609 ACCCTCTCTTTTAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668
Qy 662 ttttggagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 721
Db 669 TTTTCAAGGCTTGTGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 728
Qy 722 attgacccatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 780
Db 729 ATTGGCCCTGATGTTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 788
Qy 781 gatgaaagtttctcagagaacatgcttcttcttcttcttcttcttcttcttcttcttct 839
Db 789 GATGAGGATTTTCAAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 848
Qy 840 actttaaataatgaggtctgacacaaagaagaagaagaagaagaagaagaagaagaaga 890
Db 849 ACTTTTAAATTTGAGGGCTGAGAC-AAARAAGAGATATGAGTTGGA 898

RESULT 9
AL521483 935 bp mRNA EST 13-FEB-2001
LOCUS AL521483 L1_NFL004_NBC2 Homo sapiens cDNA clone CS0DB01YN17 3
DEFINITION prime, mRNA sequence.
ACCESSION AL521483
VERSION AL521483.1 GI:12784976
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 935)
AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.
```

```
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
1. 935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB01YN17"
/clone_id="L1_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@life.com"
http://fulllength.invitrogen.com"
BASE COUNT 254 a 182 c 186 g 286 t 27 others
ORIGIN
Query Match 54.6%; Score 800.4; DB 105; Length 935;
Best Local Similarity 92.2%; Pred. No. 7.2e-159;
Matches 864; Conservative 14; Mismatches 55; Indels 4; Gaps 4;
Qy 216 tctcaagaactcaagaatcatcatctcaagaaggggttccatgaattaccag 275
Db 935 TCTCAAGAAAGTCAAGAAATCATTTGTCAAAGAGGCTTCCATGATTCACACAG 876
Qy 276 gtctcttgaaggtcagaagaatgctgataatcatctcaagaagctggaatgga 335
Db 875 GWTTCCTTTGAGGCTCAGAGAAATGCTGATTAACATCTCCAAAGAACTGGGAATGGA 816
Qy 336 ggaagaagatgattgaagttatcagaacaaacgggggttcatcacaagttagat 395
Db 815 GGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 757
Qy 395 attcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 454
Db 756 ATTCTTTTATTTTCTTCCCAATCTCTTTATCGCTCAAAATATGCTCTNT 697
Qy 454 tgaatgtggttcaaaaggaatgaaacatgaccccatctcttgaacaactg 514
Db 696 TGNAAAGNGGTGTCNCAAAAGGAAGGAAMCGGACCCCATCTCCCTGAACATCGCG 637
Qy 514 taattgaattcagtgctcattatcattatgttgttcttcaatgctgattttg 574
Db 636 TAATTGAATTCATGCTCATTTATTTATTTGTTTCAATGCTGATGATTTTGG 577
Qy 574 tgataagctcagtccttcatatataccttcttctttaaataagctgagacag 634
Db 576 TGACCAAGCTCAGCCCTCATATTAATACCCCTCTTAATAATACAGTGTGACAG 517
Qy 634 agaggtacacttctcagacatcatttcaagctgtgtgtgataaataagctgacc 694
Db 516 AAGAGTCAACCTTTCAGACATGACATTTTCAAGCTGTGTGATTAATGATGACAC 457
Qy 694 aatgaagtgatcattgaatcttcaatggccctgagctc-agaatgattactc 753
Db 456 AATGCAAGTTCATTAATGACCTCCCAATTTGCCCCGAGTGTTCGATGATTAATTC 397
Qy 754 actcctgagctgacttcaatgaggaaggaaggttctcagagaactgagctgaa 813
Db 396 ACCCCGAGCTGACTTTCAGTGGAGATGGAAGTTTTCAGAGAACTGAATCTGGA 337
```

OY	814	aatgaagccttctctaacttgaaagccacttttaaatattgagggtctgcacaaaagaag	873
Dd	336	AAATGACCTTTCCCTAACCTGACTACTBTATAAATTGGAGGGGTGGCACAAAAGAAG	277
OY	874	aggaatacataaggttgaagtcacaagatcacagataaaggtgagagcaatgacctaacccaag	933
Dd	276	AGGAATATCAGGTTTGAAGTCAGAGATGACAGATTAAGGTGAGTAATAGCTAATCTCCAAG	217
OY	934	atggcttcaactgaagaagaaggcattttaagattttttaaaaaactgtcgagaagtccca	993
Dd	216	ATGGCTTCACTGAGAAAAAGGCATKTTTAAGATTKTTTAAAAAATCTGTCAGAAGAACC	157
OY	994	gaaagcttcaatttcatttagcaattaataaagcattcatgcagaaatgaatatcaaca	1053
Dd	156	GAAAGTTTATATTTTCATTAGCAATTAATAAGCTACATGACAAATGAACAACACACA	97
OY	1054	gaacactgctcttttagattttatttgttactttttgcttgagataygttttaaty	1113
Dd	96	GAACTGTC-CCTTTTGATGCGCATATGTTGACSGCTNNNGCCTGGGAATGAGGCCCAAATG	38
OY	1114	gacattgtctgcacagcttcatttaataaacaaca	1150
Dd	37	GACATTGTCTGTACCAAGCTTCATTAAATAAACACATA	1
RESULT	10		
AL583408/C			
LOCUS	AL583408	933 bp	mRNA
DEFINITION	AL583408 LTI_NFL010_BC2 Homo sapiens cDNA clone CSODL012YD05 3 prime mRNA sequence.	EST	16-PEB-2001
ACCESSION	AL583408		
VERSION	AL583408.1	GI:12952341	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Ll,W.B., Gruber,C., Jesse,J. and Polayes,D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr. Location/Qualifiers		
FEATURES			
source	1..933		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CSODL012YD05"		
	/clone_lib="LTI_NFL010_BC2"		
	/sex="male"		
	/tissue_type="B cells from Burkitt lymphoma"		
	/note="Vector: PCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	312 a	166 c	149 g
ORIGIN		298 t	8 others
Query Match	53.2%	Score 779.6;	DB 106; Length 933;
Best Local Similarity	95.3%;	Pred. No. 1,8e-154;	
Matches	871; Conservative	8; Mismatches	23; Indels
			12; Gaps
			7;
502	ttgaacaatctgtgaatctgaatctcagtcattacattacattatgttgytttccatgy	561	
	:		

Db	922	TTGAMACATCTGGTAATTTGGAATCTAGGCTCAATTAATCAAT	-----ATGTTTGTTGTTCAATG	868
QY	562	tgctgatttttgatcatcaagcgtcctcaagtcctccctcatatcaaccctcctttttaa	aat	621
Db	867	TCCTGA-TTTTGGTATCAAGCGCTCAGTCCCTTCATATTAAACCTCTCTTTTAA	AAAT	809
QY	622	tacgtgtcacagagaggttcacctttttcagagacattgcatttcaagcttgtgtgata	681	
Db	808	TACGGTGTGCAGAGAGAGGTACCTTTTTCAGAGACATTTGCAATTTTCAGGCTTG	GTGATA	749
QY	682	aataagatcgacccaatgtaagtglttcataatgacttccaattggcccgtatgtc-agg	740	
Db	748	AATTAAATCGACCAATGCAAGTGTTCATTAATGAAGACTTTTCCAAATGGCCCTAT	GTCTACG	689
QY	741	atgtaattacttactctctggaactgtgacttccagtgaggatgaaagtgttttcaga	gaa	800
Db	688	ATGTATTACTTCACTCACTCTCGACTGTGACTTTCAGTGGGAGATGGAAGTTTTC	AGAGAA	629
QY	801	ctgaactgtgaaaaaatgaccttcccttaacttgaagctactctttaaatttgaaggtc	860	
Db	628	CTGAACCTGTGAAAAATATACCTTTTCTTTAACTTGAACCTACTTTTAAATTTG	AGGGTCT	569
QY	861	ggaccacaaagaagagaaatcaagtggttgaagtcagaatgacagataagtgagagta	atg	920
Db	568	GGACCAAAAGAAAGAGATATACAGTTTGAAGTCAAGATGACATFATAAGTGAG	AGATGATG	509
QY	921	actaacctccaagatlggtcttactcgaagaaagacatttgaagattttttaa	aaacttg	980
Db	508	ACTTAATCTCAAAAGATGGCTTCACTGAAAGAAAAGCATTTTAAATTTTAA	AAATCTTG	449
QY	981	tcaagaagaatcccaagaagaagttcctaatttcaatgaacataaagaactacatgcaga	1044	
Db	448	TCAGAAAGATCCACAGAAAGTTCTAATTTTCATTAGCAATTAATAAAGCTAT	ATCATGTCACA	389
QY	1041	aatgatacaacagaaacactgctctttttgatttatttgaacttttgccctggata	1100	
Db	388	AATGAATATCAACACAGAACCTGCTCTTTT-GAATTTATTTTGTACTTTTGGCC	TGGGATA	330
QY	1101	tgagttttaaaatgacatgtgctgtacagcttcatataaataaacacatgtgtca	aa	1160
Db	329	TGGGTTTAATATGACAACTGTGTCTGTACCACTTCATTTAAATATTAACA	ATRTTGT-AAAA	271
QY	1161	atcgtactaatgcttatttatttatttgaatgtatagaagaagaaaaatgcctaa	aaatgaag	1220
Db	270	ATCATACTAAAGCTTATTTTATTTTATTTGTAATAGAAAGMAAAATGAGCT	TAAATTAAGG	211
QY	1221	tttcttgacaaatactgtgaaatltgacactgtgtacaatttttcttactacgtgaca	1280	
Db	210	TTCTCTTGCATTAATTAATGCGAATATGACATGTGTACAAATTTTCTTATTA	CTGTACA	151
QY	1281	ggagatgatltaatgacttttg--gaagactgaaagactgaaagtgctcttcgaa	taacag	1338
Db	150	GTGATGATGTTAATGAAGCTGTGAAGACAACTGAAAGTACTGAAAGTCT	TBTGAATCAAG	91
QY	1339	gatttaattaaagccacaaatctttttaaactactcagtggtctgtttttttaa	aaactt	1396
Db	90	GATTTAATTAAGGGGTACAAATACCTTTTATTAATCTCAGTGTCTG-TTTT	TTTAAAACTT	32
QY	1399	gataatcccgatg	1412	
Db	31	GATATTCCTGTATG	18	
RESULT	11			
LOCUS	AK002536			
DEFINITION	AK002536	1175 bp	mRNA	08-FEB-2001
ACCESSION	AK002536			
VERSION	AK002536.1	GI:12832590		
KEYWORDS	CAP	trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male kidney cDNA, RIKEN full-length enriched library, clone:0610011F08, full insert sequence.			
	Mus musculus (strain:C57BL/6J) adult male kidney cDNA, RIKEN full-length enriched mouse cDNA library			



QY	713	gacttcccaattggcccgatgcttcagcatatgcatctgaactcctctggacgttact	772
Db	738	ACCTTTCAATTGGCCTCTAAGTCCACGATGTGGTTGCTTCACTCTGGACGTGATTTT	797
QY	773	caatgagagatgaaagtttttcagagaactgtaactgtgaaaaatgaccttccact	832
Db	798	CAGTGGAGATGGAATTTTTCAGAGAACTGAACGTGTGG-AAAATGACCTTTCCTCAGCT	856
QY	833	tgaagctacttttaaatattgagggcttggaccaaaagaaggaaatcatcagttgaact	892
Db	857	TGAAGCTACTTTTAAAACTGCGGGCTGTGGACCAAAAAGAAACATCATC--GTTTGTAGT	914
QY	893	caaatgacagataaagatgtagagatgtaactgaactccaaaagatggtcttcaactgaa	952
Db	915	CAAGATGACAGATACAGTGGAGAGCAACACTAATCTCAAAAGGTGGCTTCACTGGAGGAA	974
QY	953	ggcactttaagattttttaaanaatctgttcagaaagatccccagaaaaagttctaattcat	1012
Db	975	AGAAAGTCTCTTGAGCAAGACAGTATTTGCAAAAGATCCAGAGAAAGTCTAATGTTCAT	1034
QY	1013	tacga---attaataagctatacatgctcagaatgtaatacagaacactgctctttt	1069
Db	1035	CAGCAGTATTTATTAATAGTTACTCATACGAAGTGTATGCCAACAGA----CACTGCTTT	1089
QY	1070	agattattatgtactcttttggcctgctggaatgtagttttaaatagacatgtctgacca	1129
Db	1090	TGATCTCTCTTTGCTTACTTTTGGCTGGGACATAGGTGTTTCAAGGGACATCGTCTGTACCA	1149
QY	1130	gcttcataataataacaatatctgt	1155
Db	1150	GCTTCATTTAAATTAACATATTTGT	1175
RESULT 12			
AK011074			
LOCUS	AK011074	1180 bp	mRNA
DEFINITION			HTC 08-FEB-2001
			Mus musculus 13 days embryo liver cDNA, RIKEN full-length enriched
			library, clone:2510040N03, full insert sequence.
ACCESSION			
VERSION			AK011074
KEYWORDS			AK011074.1 GI:12846958
SOURCE			CAP trapper.
			Mus musculus (strain:C57BL/6J) 13 days embryo liver cDNA to mRNA,
			clone.lib:RIKEN full-length enriched mouse cDNA library
			clone:2510040N03.
ORGANISM			Mus musculus
REFERENCE			Eukaryota; Metazoa; Chordata; Echinodermata; Vertebrata; Euteleostomi;
AUTHORS			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL			Carninci, P. and Hayashizaki, Y.
TITLE			High-efficiency full-length cDNA cloning
REFERENCE			Methods Enzymol. 303, 19-44 (1999)
AUTHORS			2 (sites)
			Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
			Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
			Normalization and subtraction of cap-trapper-selected cDNAs to
			prepare full-length cDNA libraries for rapid discovery of new genes
			Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL			20499374
MEDLINE			
REFERENCE			3 (sites)
AUTHORS			Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
			Kono, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
			Kiuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T.,
			Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T.,
			Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
			Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
			Matsura, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.
TITLE			RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL			sequencing pipeline with 384 multicapillary sequencer
MEDLINE			Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE			20530913
AUTHORS			4 (sites)
			The RIKEN Genome Exploration Research Group Phase II Team and
			FANTOM Consortium.

[illegible]

```

|||||
Db 88 GACCCCGGCTAATCCA-CGTACCATGTCTGACGAGAGCAAAACCTTCAACTGAG 146
Oy 122 acttgggggataagaagaagtgatataataactaaagtcatttggacaagataca 181
|||
Db 147 ACTTAGGCGATTAAGAGAGAGGAAATACATTAACCTCAAAATTTTGTGACGAGTACCA 206
Oy 182 gtagatcattcaaatgaaatgaaatgaaacacatcctcaagaactcaagaatcatt 241
|||
Db 207 GTGAGTACATTTCAAGATGAAATGACACACATCTCAAGAACTCAAGAAATCATACT 266
Oy 242 gtcaagaacagggtgttcccaatgaaatcactcaagttctcttgaagggtcaagaattg 301
|||
Db 267 GTCAAGAGACAGGAGATCCATGAATTCCTCAGGTTCTCTTTGAAGGTCAAGAAATTTG 326
Oy 302 ctgataatcattcctcaaaagaactcgtggaatgagaagaatgattgattgattc 361
|||
Db 327 CTGATATCATACTCGAAGAACTGGGATGAGAGAGAGATGATGATGAAGTTATC 386
Oy 362 aggaacaacagggtgtcattcaacagttatgata-----ttcttttatttttttctt 416
|||
Db 387 AGAACAACAGGGGGGCTCAGCTGAGGTTAGATATCTTTTATTTTATTTTCT 446
Oy 417 ttccctcaatccttttattttaaataagttcttgaattgtggtgttcaaacg- 475
|||
Db 447 TTTCCCTCAATCTTTTATTTTAAATAGTCTTTTGTATGTGTGTCATAAATA 506
Oy 476 gaattgaacatgagaccatctcttgaacatcgtgaattgaattcattgattc 535
|||
Db 507 AAATGTGAATCGACCTCATCTTTAGAA-----CATATGATTTCTAGTTTCA 557
Oy 536 ttattcaatttgttcttcttcaatgttctgatttggatgaagctcagctccct 595
|||
Db 558 ATATTCATTTTGTGTTGTTTGTGTCGATGATTTTGTGATCAGACCTCAGCTCCCT 617
Oy 596 catactacccctccttttaa---aataagtgagcagaagaagtgacccctttcag 652
|||
Db 618 AATTCCTCTTTTCTTTTAAAGATTTTCAATGATGTCACAGAGAGGCCACCTTTTCA 677
Oy 653 gaattcatttcaagctgtgtgtgataaataagatcgacacatgcaagtgttcatat 712
|||
Db 678 GACTGTGCTATTTTCAAGTTTGTGATGATTA--AAGATCGACCAATGGAGCTTTCTATG 736
Oy 713 gacttccaattgacctgattgttcaagatgttacttcaactccctggaattgactt 772
|||
Db 737 ACCTTTCATTTGGCTCTGAGATTCACAGATGTGTTCTTCTCAGCTGAGCTGATTTT 796
Oy 773 caatgggagatggaagtttccagagaactgaactgttgaagaatgaccccttcaact 832
|||
Db 797 CAGTGGGATGGAATTTTTCAGAGAACTGAATCTGTG--AAATGACCTTTCTCAGCT 855
Oy 833 tgaagctactttaaatttgaaggtctgagcacaagaagaagaatatacaagttgaagt 892
|||
Db 856 TGAAGCTACTTTTAAATCTCGGGTCTGACCAAAAGAAACATCAT--GTTTGTAGT 913
Oy 893 caagatgacagataagtgagatgaatgaactccaagaatgtgttcaactgaagaaga 952
|||
Db 914 CAAAGTGAACATACAGAGAGCAACAGCTTACCTCCAAAGGTGCTTCACTGGAGACA 973
Oy 953 ggcatttaagaattttaaataatcttgcagaagaatcccgaaagaatttcaatttcat 1012
|||
Db 974 ACAAAGTGTCTTGAGCAGACAGATTTCTCAGAAAGATCCAGGAAGTTTAAATGTTAT 1033
Oy 1013 tagca---attaataaagctataacagagaatgaatgaacaagaacacgtcctttt 1069
|||
Db 1034 CAGCAGTATTTAATAAGTTACTCATACAGAGTGTACGCAACAA-----CACTGCTTT 1088
Oy 1070 agatttatttgaatttggccgtgagataggttttaaatgacattgtctgacca 1129
|||
Db 1089 TCATTTCTGTTTACTTTTGGCTGGGACATGGGTTTCAAGGACATCGTCTGTACCA 1148
Oy 1130 gcttcaataaataaacaatttgcaca 1159
|||||

```

```

Db 1149 GCTTCAATTAATAACATATTTGTAAAA 1178

RESULT 13
AL532393/c 943 bp mRNA EST 13-FEB-2001
LOCUS AL532393 LTL_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014YH24 3
DEFINITION prime, mRNA sequence.
ACCESSION AL532393
VERSION AL532393.1 GI:12795886
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 943)
AUTHORS Li, W.-B., Gruber, C., Jassie, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segrel@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..943
location/Qualifiers
/oranism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM014YH24"
/clone_id="LTL_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCMVSPORT 6
vector. Library was normalized. Library was constructed
by life Technologies. Contact : Feng Liang Life
Technologies, a division of invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@life.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 199 a 185 c 187 g 294 t 78 others
ORIGIN

Query Match 50.3%; Score 737; DB 106; Length 943;
Best Local Similarity 87.0%; Pred No. 1.7e-145;
Matches 824; Conservative 6; Mismatches 111; Indels 6; Gaps 6;

Oy 184 gagattcaactcaagtgaaatgaaacacacatcgaagaactcaagaatcaactactgt 243
|||
Db 943 GAGATTCACATCAAGGAAATGACACACATCTCAAGAACTCAAGAAATCATCTGT 884
Oy 244 caaagacagggtgttccaatgaattcaatttcttggagggtgaagaattgtc 303
|||
Db 883 CAAAGACAGGGTGTCCCAATGAATMCATCAGGTGCTCTTTGAGGTCAGAAATTTGCT 824
Oy 304 gataatcatatctc--aaagaactcgggaatggaagaagaatgattgaattatca 362
|||
Db 823 GATATATATCTCCAAAAGAACTGGAATGAGAGAAAGATGTGATTCACAAWTTATCA 764
Oy 363 ggaacaacagggtgtcattcaacagtttagatatttcttatttcttcttccct 422
|||
Db 763 GGAACAACAG--GGGGKCATTTCAACAGTTAGATATCTTTTATTTTCTTTTCCCT 705
Oy 423 caatccttttattttaaataagttcttggtaatggtgtgttcaaaaggaattga 482
|||
Db 704 CAATCTTTTATTTTAAATAATATGTTG--TTTGAANNGGNNNNCAAAACGAANNNGA 646
Oy 483 aaattgacacccatcttgaagaactcgtgaattgaattcattgattgtcattatca 542
|||
Db 645 AACNGGCACCCCAACNCNCNGAACAACNCNGAANNNGAANNCAANNCAANNANCA 586
|||||

```



[illegible]



[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 16:36:04 ; Search time 72.32 seconds  
(without alignments)  
28.736 Million cell updates/sec

Title: US-09-484-964-2

Sequence: 1 MSQDEAKPRFEDGDKKQGF.....MEEDYIEYQEGTGHSTV 101

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCFUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223.5	42.6	104	2	US-08-853-974-3
2	223.5	42.6	104	4	US-09-172-988-3
3	217	41.3	126	2	US-08-853-974-1
4	217	41.3	126	4	US-09-172-988-1
5	101	19.2	412	2	US-08-755-584-2
6	101	19.2	412	3	US-09-192-611-2
7	76	14.5	352	3	US-08-854-764-2
8	76	14.5	352	5	PCT-US95-09377-2
9	73	13.9	75	1	US-08-350-884-35
10	73	13.9	75	1	US-08-709-173-35
11	73	13.9	75	2	US-08-709-177-35
12	73	13.9	75	2	US-08-833-678A-3
13	73	13.9	75	4	US-08-529-169A-3
14	73	13.9	76	1	US-08-232-815-2
15	73	13.9	76	1	US-08-350-906-2
16	73	13.9	76	2	US-09-070-060-8
17	73	13.9	76	3	US-09-357-746-8
18	73	13.9	76	5	PCT-US95-04536-2
19	73	13.9	78	3	US-08-505-486-94
20	73	13.9	78	3	US-08-801-028-94
21	73	13.9	78	3	US-09-340-154-94
22	73	13.9	78	5	PCT-US95-09338-94
23	73	13.9	78	5	PCT-US95-09339-94
24	73	13.9	103	2	US-08-771-201-9
25	73	13.9	114	2	US-08-771-201-10
26	73	13.9	128	2	US-08-557-309B-41
27	73	13.9	128	3	US-08-834-306-41

28	73	13.9	128	4	US-08-993-674A-41	Sequence 41, Appl
29	73	13.9	147	2	US-08-771-201-11	Sequence 11, Appl
30	73	13.9	156	2	US-09-070-060-7	Sequence 7, Appl
31	73	13.9	156	3	US-09-051-969A-3	Sequence 3, Appl
32	73	13.9	156	3	US-09-051-969A-4	Sequence 4, Appl
33	73	13.9	156	3	US-09-357-746-7	Sequence 7, Appl
34	73	13.9	229	2	US-08-726-306A-23	Sequence 23, Appl
35	73	13.9	229	3	US-08-840-146-20	Sequence 20, Appl
36	73	13.9	229	3	US-09-360-220-20	Sequence 20, Appl
37	73	13.9	533	6	5510474-2	Patent No. 5510474
38	73	13.9	1121	1	US-07-789-915A-2	Sequence 2, Appl
39	73	13.9	1121	1	US-08-005-002C-2	Sequence 2, Appl
40	73	13.9	1121	1	US-08-487-203A-2	Sequence 2, Appl
41	71	13.5	76	1	US-08-450-834-6	Sequence 6, Appl
42	71	13.5	77	6	5510474-3	Patent No. 5510474
43	71	13.5	533	1	US-08-462-092-2	Sequence 2, Appl
44	71	13.5	533	3	US-08-746-822-2	Sequence 2, Appl
45	71	13.5	533	3	US-09-094-350-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1

US-08-853-974-3

Sequence 3, Application US/08853974

Patent No. 5840534

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Shah, Purvi

TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTED for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/853,974

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0289 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 104 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: Genbank

CLONE: 1707372

US-08-853-974-3

Query Match 42.6% ; Score 223.5 ; DB 2 ; Length 104 ;  
Best Local Similarity 51.2% ; Pred. NO. 4,2e-19 ;  
Matches 44 ; Conservative 14 ; Mismatches 25 ; Indels 3 ; Gaps 1 ;



```

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/172.988
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/853.974
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PR-0289 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGEPT03
CLONE: 2361410
US-09-172-988-1

```

```

41.3% Score 217; DB 4; Length 126;
Query Match Similarity 44.9%; Pred. No. 3,1e-18;
Best Local Similarity 44.9%; Pred. No. 3,1e-18;
Matches 44; Conservative 20; Mismatches 28; Indels 6; Gaps 2.

OY 1 MSDOEAKPSTEDIDGDKKQGEYIKLIVIQDSSSEIHFKVKTTHLKLKESYCOROGVPM 60
    ||::: | - - - - - | - - - - - | - - - - - | - - - - - | - - - - - |
Db 1 MSSEKPE-----GVKTENDHINLVAQDSVQFKIKRRTPLSKLMKVCRRGSLNR 55
    :||::| | | : | : | | | | | : | | : | | : | | : | | : | | : | |

OY 61 SLRFEGORIADNTPKELGMEEDYEVY-QEQTG 97
    :||::| | | : | : | | | | | : | | : | | : | | : | | : | | : | |
Db 56 QIRRFEDGQPINETDTPAOLEMEDEDITDVPSSKTG 93

RESULT 5
US-08-755-584-2
; Sequence 2, Application US/08755584
; Patent No. 5858711
; GENERAL INFORMATION:
; APPLICANT: Glimmer, Laurie H.
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUI-026

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; OS-08-755-584-2

```

```

Query Match      19.2% Score 101; DB 2; Length 412;
Best Local Similarity 25.5%; Pred. No. 0.00057;
Matches    26; Conservative   22; Mismatches   38; Indels   16; Gaps   2.

QY      6 AKPSTEDLG-----DKKGGYIKLKVIGSDSS--IHKVMTHTLKIKE 49
          ||| |
DB     310 ATPSTLTGLGADIIDCVVLASSSEATETTSQELRLRVCKGEKHQMILISTSPDSPFLVKLS 369
          ||| |
QY     50 SCYCGGVPMNSLRPLFEGQRADNHPKEKLMEDEEITYEV 91
          | : : : : : | : : : : : | : : : : : | : : : : :
DB     370 HYEPAMGLSGHKLSTFFFDGTKLSCGELPADLDGLESBGDILEWV 411

```

```

1      RESULT      6
2      US-09-192-611-2
3      ; Sequence 2, Application US/09192611
4      ; Patent No. 6090561
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Glimcher, Laurie H.
7      ; APPLICANT: Hodge, Martin R.
8      ; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
9      ; TITLE OF INVENTION: OF USE THEREFOR
10     ; NUMBER OF SEQUENCES: 2
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: LAHIVE & COCKFIELD
13     ; STREET: 60 State street, suite 510
14     ; City: Boston
15     ; STATE: Massachusetts
16     ; COUNTRY: USA
17     ; ZIP: 02109-1875
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/09/192,611
25     ; FILING DATE:
26     ; CLASSIFICATION:
27     ; PRIOR APPLICATION DATA:
28     ; APPLICATION NUMBER: US/08/755,584
29     ; FILING DATE:
30     ; ATTORNEY/AGENT INFORMATION:
31     ; NAME: Giulio A. DeConti, Jr.
32     ; REGISTRATION NUMBER: 31,503
33     ; REFERENCE/DOCKET NUMBER: HUI-026
34     ; TELECOMMUNICATION INFORMATION:
35     ; TELEPHONE: (617)227-7400
36     ; TELEFAX: (617)227-5941
37     ; INFORMATION FOR SEQ ID NO: 2:
38     ; SEQUENCE CHARACTERISTICS:
39     ; LENGTH: 412 amino acids
40     ; TYPE: amino acid
41     ; TOPOLOGY: linear
42     ; MOLECULE TYPE: protein
43     ;
44     ; US-09-192-611-2
45
46     Query Match      19.2%; Score 101; DB 3; Length 412;
47     Best Local Similarity 25.5%; Pred. NO. 0.00057;
48     Matches 26; Conservative 22; Mismatches 38; Indels 16; Gaps 2

```



\_\_\_\_\_

100

OY 94 QTGG 97  
Db 72 LRGG 75

## RESULT 12

US-08-833-678A-3  
; Sequence 3, Application US/08833678A  
; Patent No. 5989905

## GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL  
APPLICANT: CHOO, OUI-LIM

APPLICANT: HAN, JANG

APPLICANT: CHOE, JOONHO  
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING  
TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION  
STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville  
STATE: California

COUNTRY: U.S.A.  
ZIP: 94662-8097

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/833.678A  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/529,169  
FILING DATE: 15-SEP-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895  
REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 923-3274

TELEFAX: (510) 655-3542  
TELEX: n/a

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

LENGTH: 75 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-833-678A-3

Query Match 13.9%; Score 73; DB 2; Length 75;  
Best Local Similarity 20.3%; Pred. No. 0.12;  
Matches 13; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

OY 34 IHFKYMTHTLKKESYQROGVPMNSLRFLFEGORIDNHTPKELGMEEDVLEVYOE 93  
Db 12 ILEVESSDTIDNVKSKIDKGIIPDQORLLIFAGKQLEDGRTLSYNIQKESTLHLVLR 71

OY 94 QTGG 97  
Db 72 LRGG 75

## RESULT 13

US-08-529-169A-3  
; Sequence 3, Application US/08529169A  
; Patent No. 6194140

## GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL

APPLICANT: CHOO, OUI-LIM  
APPLICANT: HAN, JANG

APPLICANT: CHOE, JOONHO  
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING  
TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION  
STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville  
STATE: California

COUNTRY: U.S.A.  
ZIP: 94662-8097

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/529,169A  
FILING DATE: 15-SEP-1995

CLASSIFICATION: 4325  
ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.  
REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0100.005  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-3274  
TELEFAX: (510) 655-3542

TELEX: n/a  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-529-169A-3

Query Match 13.9%; Score 73; DB 4; Length 75;  
Best Local Similarity 20.3%; Pred. No. 0.12;  
Matches 13; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

OY 34 IHFKYMTHTLKKESYQROGVPMNSLRFLFEGORIDNHTPKELGMEEDVLEVYOE 93  
Db 12 ILEVESSDTIDNVKSKIDKGIIPDQORLLIFAGKQLEDGRTLSYNIQKESTLHLVLR 71

OY 94 QTGG 97  
Db 72 LRGG 75

## RESULT 14

US-08-232-815-2  
; Sequence 2, Application US/08232815  
; Patent No. 5503977

## GENERAL INFORMATION:

APPLICANT: Johnson, Nils  
APPLICANT: Varshavsky, Alexander

TITLE OF INVENTION: SPLIT UBIQUITIN PROTEIN SENSOR  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Kevin M. Farrell  
STREET: P.O. Box 999

CITY: York Harbor  
STATE: Maine

COUNTRY: USA  
ZIP: 03911

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible



US-08-350-906-2

Query Match	13.9%	Score 73;	DB 1;	Length 76;
Best Local Similarity	20.3%	Pred. No. 0.12;		
Matches 13;	Conservative 19;	Mismatches 32;	Indels 0;	Gaps 0;

OY 34 IHKRVMTTHLKKLKESYCQROGVPMNSLRFLTEGORADNHPRKEJGMEDEEVYEOE 93  
| : : : : | : : : : | : : : : | : : : : | : : : :  
Db 13 ILEVESSPTIDWVKSIQDKREGIPPDQGRLIFMAGKLQEDSRTLSVDNIKESTIHLVLR 72

OY 94 QTGS 97  
||  
Db 73 LRGS 76

```
Search completed: August 15, 2001, 17:15:28
Job time: 2364 sec
```

Query Match	13.9%	Score 73;	DB 1;	Length 76;
Best Local Similarity	20.3%;	Pred. No. 0.12;		
Matches 13; Conservative	19;	Mismatches 32;	Indels 0;	Gaps 0;

Dy 34 IHRVKMTFLKLKLESYCQROGVPMSNLRFLEGGRIADNHTPKELGNEEDVIEYQE 93  
| : | : : | : : | : | : : | : : : | : : : :  
Db 13 TLEVESSDIDNVKSKIQDKESIPPDORLIFAGQLDEGRITSDYNIQKESTLLVLVR 72

QY	94	QTGG	97
Db	73	LRGG	76

RESULT 15  
US-08-350-

US-08-350-906-2  
; Sequence 2, Application US/08350906  
; Patent No. 5585245

GENERAL INFORMATION:  
APPLICANT: Johnsson, Nils  
APPLICANT: Varshavsky, Alexander  
TITLE OF INVENTION: UBIQUITIN-BASED SPLIT PROTEIN SENSOR  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kevin M. Farrell  
STREET: P O Box 999  
CITY: York Harbor  
STATE: Maine  
COUNTRY: USA

ZLP: 0391

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: IBM PC COMPATIBLE

SOFTWARE: Patent In Release #10 Version #1 25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/350,906

FILING DATE: 07-DEC-1994

CLASSIFICATION: 435  
ATTORNEY/AGENT INTRODUCTION:NAME: Farrell, Kenneth  
AI ORNER/AGENT INFORMATION:

NAME: FARRILL, KEVIN M.  
REGISTRATION NUMBER: 35 505

REFERENCE/DOCKET NUMBER: CIT-9303A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 207 363-05

TELEFAX: 207 363-0528

```

; INFORMATION FOR SEQ ID NO: 2
SEQUENCE CANNOT BE DISPLAYED

```

SEQUENCE CHARACTERISTICS: 76 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

Thu Aug 16 08:48:53 2001

us-09-484-964-2.rai

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 16:16:34 ; Search time 63.2 Seconds  
(without alignments)  
96.883 Million cell updates/sec

Title: US-09-484-964-2  
Sequence: 1 MSDEAKPSTEDLGKKQGE.....MEEDVLEVYQEQTGHSTV 101

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues  
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSI/gcgdata/geneseq/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525	100.0	101	19	AAW60079 Homo sapiens sentr
2	522	99.4	101	20	AAW60079 Homo sapiens sentr
3	522	99.4	101	21	AAW60079 Homo sapiens sentr
4	522	99.4	101	21	AAW60079 Homo sapiens sentr
5	239.5	45.6	95	20	AAW60079 Homo sapiens sentr
6	232	44.2	95	20	AAW60079 Homo sapiens sentr
7	223.5	42.6	100	21	AAW60079 Homo sapiens sentr
8	221	42.1	103	21	AAW60079 Homo sapiens sentr
9	221	42.1	103	21	AAW60079 Homo sapiens sentr
10	221	42.1	153	21	AAW60079 Homo sapiens sentr
11	218.5	41.6	100	20	AAW60079 Homo sapiens sentr

12	218	41.5	99	21	AAW60079
13	217	41.3	126	20	AAW60079
14	175.5	33.4	111	21	AAW60079
15	175.5	33.4	145	21	AAW60079
16	147	28.0	83	21	AAW60079
17	147	28.0	83	21	AAW60079
18	147	28.0	108	21	AAW60079
19	147	28.0	108	21	AAW60079
20	147	28.0	142	21	AAW60079
21	147	28.0	143	21	AAW60079
22	133	25.3	57	21	AAW60079
23	127	24.2	61	21	AAW60079
24	111	21.1	66	21	AAW60079
25	104	19.8	122	21	AAW60079
26	104	19.8	215	21	AAW60079
27	104	19.8	244	21	AAW60079
28	101	19.2	153	21	AAW60079
29	101	19.2	190	21	AAW60079
30	101	19.2	412	18	AAW60079
31	101	19.2	412	20	AAW60079
32	100	19.0	419	20	AAW60079
33	98	18.7	154	21	AAW60079
34	98	18.7	156	21	AAW60079
35	98	18.7	169	21	AAW60079
36	96	18.3	78	21	AAW60079
37	94.5	18.0	153	22	AAW60079
38	94.5	18.0	183	22	AAW60079
39	94.5	18.0	220	22	AAW60079
40	94.5	18.0	250	22	AAW60079
41	94.5	18.0	262	22	AAW60079
42	94.5	18.0	296	22	AAW60079
43	94.5	18.0	451	22	AAW60079
44	94.5	18.0	464	22	AAW60079
45	94.5	18.0	464	22	AAW60079

## ALIGNMENTS

RESULT 1  
ID AAW60079 standard; Protein: 101 AA.  
AC AAW60079;  
DT 11-SEP-1998 (first entry)  
DE Homo sapiens sentrin-1 polypeptide.  
KW sentrin-1; protection; tumour necrosis factor; tnfr; apoptosis;  
KW Fas/Apo-induced; tumour cell death; induction; tumour aggressiveness;  
KW detection; determination.  
XX Homo sapiens.  
XX WO9820038-A1.  
XX 14-MAY-1998.  
XX 05-NOV-1997; 97WO-US20344.  
XX 05-NOV-1996; 96US-0030302.  
XX (TEXA) UNIV TEXAS SYSTEM.  
XX Yeh ETH;  
XX WPI: 1998-286668/25.  
XX N-PSDB; AAV34564.  
XX New isolated sentrin polypeptide(s) - which inhibit TNF receptor or  
XX Fas/Apo-induced apoptosis, used to develop products for inducing  
XX cell death in tumours

XX Claim 4; Page 78; 120pp; English.

PS The sequence is that of the sentrin-1 polypeptide.  
 CC Sentrin polypeptides have the ability to protect or guard  
 CC cells from tumour necrosis factor (TNF) or Fas/APO induced  
 CC cell death (apoptosis). Inhibitors of the sentrin polypeptides,  
 CC e.g. antibodies, can be used for inducing cell death,  
 CC particularly in tumours. The products can also be used for  
 CC determining the aggressiveness of a tumour and for detection and  
 CC isolation of products. The sentrin polypeptide can also be used to  
 CC detect a ubiquitin conjugating enzyme polypeptide or PML polypeptide.  
 XX Sequence 101 AA;

Query Match 100.0%; Score 525; DB 19; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-54;  
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGODSSEIHFKVMTTHLKLKESYCOROGVPMN 60  
 DB 1 msdgaekpstedldgdkkgeyiklvigdsseihfkvmtthlklkkesycqrgvpmn 60  
 QY 61 SIRFLFEGORLADNHTPKELGMEEDVIEVYQGTGSHSTV 101  
 DB 61 sirflfeggrladihntpkelgmeedvievyqgtgshstv 101

## RESULT 2

AAW87984  
 ID AAW87984 standard; peptide; 101 AA.

XX AAW87984;  
 AC  
 XX 29-APR-1999 (first entry)

DE Ubiquitin-like domain of the protein SUMO1.

XX Ubiquitin; ubiquitin/proteasome pathway; degradation signal; UBL domain;  
 KW intracellular protein degradation; proteasome; ubiquitin-like domain;  
 KW fusion protein; drug resistance; malignant cell; purification;  
 KM SUMO1.

XX Mammalia.

XX WO9857978-A1.

XX 23-DEC-1998.

XX 19-JUN-1998; 98WO-US12846.

XX 19-JUN-1997; 97US-0050171.

XX (UYNE-) UNIV NEW JERSEY.

XX Madura K;

XX WPI; 1999-095322/08.

XX Ubiquitin-like domains as fusion or immobilised proteins - useful,  
 PT respectively, for assessing the proliferative potential of malignant  
 PT cells or to purify proteasome complexes

XX Claim 2; Page 25; 84pp; English.

XX The present sequence represents a ubiquitin-like domain of a  
 CC mammalian protein. The ubiquitin/proteasome pathway is a major  
 CC pathway for the intracellular degradation of proteins, where the  
 CC ubiquitin acts as a degradation signal. Proteins which bind the  
 CC proteasome (a multisubunit complex) have ubiquitin-like (UBL)  
 CC domains. When the UBL domains are fused to a reporter protein,  
 CC they can be used to assess the proliferative potential and select

CC drug resistance of malignant cells. The UBL domain also allows the  
 CC purification of proteasomes for the isolation and characterisation  
 CC of novel subunits of the proteasome.

XX Sequence 101 AA;

Query Match 99.4%; Score 522; DB 20; Length 101;  
 Best Local Similarity 99.0%; Pred. No. 2.6e-54;  
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGODSSEIHFKVMTTHLKLKESYCOROGVPMN 60  
 DB 1 msdgaekpstedldgdkkgeyiklvigdsseihfkvmtthlklkkesycqrgvpmn 60  
 QY 61 SIRFLFEGORLADNHTPKELGMEEDVIEVYQGTGSHSTV 101  
 DB 61 sirflfeggrladihntpkelgmeedvievyqgtgshstv 101

## RESULT 3

AAV49967  
 ID AAV49967 standard; Protein; 101 AA.

XX AAV49967;  
 AC  
 XX 03-FEB-2000 (first entry)

DE Human sentrin protein sequence.

XX Human; sentrin; antisense oligonucleotide; phosphorothioate;  
 KW inhibition; modulation; expression; diagnosis.

XX Homo sapiens.

XX US5985664-A.

XX 16-NOV-1999.

XX 17-DEC-1998; 98US-0213768.

XX 17-DEC-1998; 98US-0213768.

XX (ISIS-) ISIS PHARM INC.

XX Baker BF, Cowsett LM;

XX WPI; 2000-022284/02.

XX N-PSDB; AA235859.

XX Antisense compound which modulates human sentrin expression, useful for  
 PT treating diseases associated with sentrin expression -  
 PT Example 13; Column 39-42; 29pp; English.

XX The present invention describes an antisense compound (I) 8-30  
 CC nucleotides long targeted to a nucleic acid molecule encoding human  
 CC sentrin. The antisense compound comprises a phosphorothioate antisense  
 CC oligonucleotide which inhibits expression of human sentrin. (I) is  
 CC useful for inhibiting expression of sentrin in human cells or tissues  
 CC in vitro, for treating humans or other animals suspected of having or  
 CC being prone to a disease associated with sentrin expression. (I) can  
 CC also be used for research or diagnostic purposes. The present  
 CC sequence represents human sentrin.

XX Sequence 101 AA;

Query Match 99.4%; Score 522; DB 21; Length 101;  
 Best Local Similarity 99.0%; Pred. No. 2.6e-54;  
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGODSSEIHFKVMTTHLKLKESYCOROGVPMN 60

```

Db      1 mdsqgkpsdtdlgdkkeeyiklvigqdsstlnrkvmtnlkkllksyccqrgvpmn 60
OY      61 SLRFLFEGORIDNHTPKELGMEEDVLEVYOQOTGSHSTV 101
        |||
Db      61 slrlfeggrldnhtpkelgmeedvlevygeqgynstlv 101

RESULT  4
AAW74776
ID      AAW74776 standard; Protein; 102 AA.
XX      AAW74776;
AC      AAW74776;
XX      25-JAN-1999 (first entry)
DT      Human secreted protein encoded by gene 47 clone HOGAV75.
XX      Human; secreted protein; testis; tumour; foetal brain tissue;
KW      fusion protein; cancer; central nervous system; seizure;
KM      diagnosis; neurodegenerative disease.
XX      Homo sapiens.
XX      Key Location/Qualifiers
FH      Misc-difference 102
FT      label= unknown
XX      MO9839448-A2.
XX      11-SEP-1998.
XX      06-MAR-1998; 98MO-US04493.
XX      02-OCT-1997; 97US-0061060.
XX      07-MAR-1997; 97US-0038621.
XX      07-MAR-1997; 97US-0040161.
XX      07-MAR-1997; 97US-0040162.
XX      07-MAR-1997; 97US-0040163.
XX      07-MAR-1997; 97US-0040333.
XX      07-MAR-1997; 97US-0040334.
XX      07-MAR-1997; 97US-0040336.
XX      07-MAR-1997; 97US-0040626.
XX      11-APR-1997; 97US-0043311.
XX      11-APR-1997; 97US-0043312.
XX      11-APR-1997; 97US-0043313.
XX      11-APR-1997; 97US-0043314.
XX      11-APR-1997; 97US-0043356.
XX      11-APR-1997; 97US-0043359.
XX      11-APR-1997; 97US-0043576.
XX      11-APR-1997; 97US-0043578.
XX      11-APR-1997; 97US-0043580.
XX      11-APR-1997; 97US-0043669.
XX      11-APR-1997; 97US-0043670.
XX      11-APR-1997; 97US-0043671.
XX      11-APR-1997; 97US-0043672.
XX      11-APR-1997; 97US-0043674.
XX      11-APR-1997; 97US-0043674.
XX      23-MAY-1997; 97US-0047500.
XX      23-MAY-1997; 97US-0047501.
XX      23-MAY-1997; 97US-0047502.
XX      23-MAY-1997; 97US-0047503.
XX      23-MAY-1997; 97US-0047581.
XX      23-MAY-1997; 97US-0047582.
XX      23-MAY-1997; 97US-0047583.
XX      23-MAY-1997; 97US-0047584.
XX      23-MAY-1997; 97US-0047585.
XX      23-MAY-1997; 97US-0047586.
XX      23-MAY-1997; 97US-0047587.
XX      23-MAY-1997; 97US-0047588.
XX      23-MAY-1997; 97US-0047589.
XX      23-MAY-1997; 97US-0047590.
XX      23-MAY-1997; 97US-0047592.

```

```

PR      23-MAY-1997; 97US-0047593.
PR      23-MAY-1997; 97US-0047594.
PR      23-MAY-1997; 97US-0047595.
PR      23-MAY-1997; 97US-0047596.
PR      23-MAY-1997; 97US-0047597.
PR      23-MAY-1997; 97US-0047598.
PR      23-MAY-1997; 97US-0047599.
PR      23-MAY-1997; 97US-0047600.
PR      23-MAY-1997; 97US-0047601.
PR      23-MAY-1997; 97US-0047612.
PR      23-MAY-1997; 97US-0047613.
PR      23-MAY-1997; 97US-0047614.
PR      23-MAY-1997; 97US-0047615.
PR      23-MAY-1997; 97US-0047617.
PR      23-MAY-1997; 97US-0047618.
PR      23-MAY-1997; 97US-0047632.
PR      23-MAY-1997; 97US-0047633.
PR      06-JUN-1997; 97US-0048964.
PR      06-JUN-1997; 97US-0048974.
PR      13-JUN-1997; 97US-0049610.
PR      08-JUL-1997; 97US-0051926.
PR      16-JUL-1997; 97US-0052874.
PR      18-AUG-1997; 97US-0055724.
PR      22-AUG-1997; 97US-0056630.
PR      22-AUG-1997; 97US-0056631.
PR      22-AUG-1997; 97US-0056632.
PR      22-AUG-1997; 97US-0056636.
PR      22-AUG-1997; 97US-0056637.
PR      22-AUG-1997; 97US-0056662.
PR      22-AUG-1997; 97US-0056664.
PR      22-AUG-1997; 97US-0056845.
PR      22-AUG-1997; 97US-0056862.
PR      22-AUG-1997; 97US-0056864.
PR      22-AUG-1997; 97US-0056872.
PR      22-AUG-1997; 97US-0056874.
PR      22-AUG-1997; 97US-0056875.
PR      22-AUG-1997; 97US-0056876.
PR      22-AUG-1997; 97US-0056877.
PR      22-AUG-1997; 97US-0056878.
PR      22-AUG-1997; 97US-0056880.
PR      22-AUG-1997; 97US-0056881.
PR      22-AUG-1997; 97US-0056882.
PR      22-AUG-1997; 97US-0056884.
PR      22-AUG-1997; 97US-0056886.
PR      22-AUG-1997; 97US-0056887.
PR      22-AUG-1997; 97US-0056888.
PR      22-AUG-1997; 97US-0056889.
PR      22-AUG-1997; 97US-0056892.
PR      22-AUG-1997; 97US-0056893.
PR      22-AUG-1997; 97US-0056894.
PR      22-AUG-1997; 97US-0056903.
PR      22-AUG-1997; 97US-0056909.
PR      22-AUG-1997; 97US-0056910.
PR      22-AUG-1997; 97US-0056911.
PR      05-SEP-1997; 97US-0057650.
PR      05-SEP-1997; 97US-0057669.
PR      05-SEP-1997; 97US-0057761.
PR      12-SEP-1997; 97US-0058785.

(PHUA-) HUMAN GENOME SCI INC.
XX      Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI      Peng P, Petrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI      Ryaw H, Latleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI      Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX      WPI: 1998-506364/43.
DR      N-PSDB; AAV59557.
XX      New isolated human genes and the secreted polypeptide(s) they encode
PT      - useful for diagnosis and treatment of e.g. cancers, neurological

```

PT disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS  
 XX Claim 1; Page 566-567; 721pp; English.

CC This sequence represents a secreted human protein encoded by the nucleic  
 CC acid molecule designated Gene 47 from the human cDNA clone HOGAV75  
 CC (deposited as clone ATCC 97899 and ATCC 209045).  
 CC The gene can be used to generate fusion proteins by linking to the gene  
 CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the  
 CC stability of the fused protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: AAV59511-595812; amino acid sequences AAW74731-W75026)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological  
 CC polypeptides can be diagnosed by determining the amount of the new  
 CC polypeptides in a sample or by determining the presence of mutations in  
 CC the new polynucleotides. Specific uses are described for each of the 186  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAV59511 for described uses).

XX Sequence 102 AA;

Query Match 99.4%; Score 522; DB 19; Length 102;  
 Best Local Similarity 99.0%; Pred. No. 2.7e-54;  
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGDSSSEIFKVKMTHTLKKESYQGRGVPNN 60  
 DB 1 msdgaekpstedldgkkggeyiklvigdseslhfkvkmthtklkesycqrgvpmn 60  
 OY 61 SLRFLFEGQRIADNHTPKRELGMEEEDVIEVYQEOGTGHSTV 101  
 DB 61 slrflfegqriadnhtpkrelgmeeedvievyyqegtghstv 101

RESULT 5

AAW87986 standard; peptide: 95 AA.

AAW87986;

29-APR-1999 (first entry)

Ubiquitin-like domain of the yeast protein SMT3A.

Ubiquitin; ubiquitin/proteasome pathway; degradation signal; UBL domain;  
 intracellular protein degradation; proteasome; ubiquitin-like domain;  
 fusion protein; drug resistance; malignant cell; purification;  
 yeast; SMT3A.

Saccharomyces sp.

MO9857978-A1.

23-DEC-1998.

19-JUN-1998; 98WO-US12846.

19-JUN-1997; 97US-0050171.

(UYNE-) UNIV NEW JERSEY.

Madura K;

WPI; 1999-095322/08.

Ubiquitin-like domains as fusion or immobilised proteins - useful,  
 respectively, for assessing the proliferative potential of malignant  
 cells or to purify proteasome complexes  
 Claim 2; Page 25; 84pp; English.

CC The present sequence represents a ubiquitin-like domain of the  
 CC yeast protein SMT3A. The ubiquitin/proteasome pathway is a major  
 CC pathway for the intracellular degradation of proteins, where the  
 CC ubiquitin acts as a degradation signal. Proteins which bind the  
 CC proteasome (a multisubunit complex) have ubiquitin-like (UBL)  
 CC domains. When the UBL domains are fused to a reporter protein,  
 CC they can be used to assess the proliferative potential and select  
 CC drug resistance of malignant cells. The UBL domain also allows the  
 CC purification of proteasomes for the isolation and characterisation  
 CC of novel subunits of the proteasome.

XX Sequence 95 AA;

Query Match 45.6%; Score 239.5; DB 20; Length 95;  
 Best Local Similarity 47.4%; Pred. No. 6.6e-21;  
 Matches 46; Conservative 20; Mismatches 26; Indels 5; Gaps 1;

OY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGDSSSEIFKVKMTHTLKKESYQGRGVPNN 60  
 DB 1 mseeakpke-----gvtendhnlkvaqdgsvqfklkrltslklmkayceerglsmr 55  
 OY 61 SLRFLFEGQRIADNHTPKRELGMEEEDVIEVYQEOGTG 97  
 DB 56 qirfrdgqpinetdtpaglmededtdvdfqgtg 92

RESULT 6

AAW87985 standard; peptide: 95 AA.

AAW87985;

29-APR-1999 (first entry)

Ubiquitin-like domain of the yeast protein SMT3B.

Ubiquitin; ubiquitin/proteasome pathway; degradation signal; UBL domain;  
 intracellular protein degradation; proteasome; ubiquitin-like domain;  
 fusion protein; drug resistance; malignant cell; purification;  
 yeast; SMT3B.

Saccharomyces sp.

MO9857978-A1.

23-DEC-1998.

19-JUN-1998; 98WO-US12846.

19-JUN-1997; 97US-0050171.

(UYNE-) UNIV NEW JERSEY.

Madura K;

WPI; 1999-095322/08.

Ubiquitin-like domains as fusion or immobilised proteins - useful,  
 respectively, for assessing the proliferative potential of malignant  
 cells or to purify proteasome complexes  
 Claim 2; Page 25; 84pp; English.

The present sequence represents a ubiquitin-like domain of the  
 CC yeast protein SMT3B. The ubiquitin/proteasome pathway is a major  
 CC pathway for the intracellular degradation of proteins, where the  
 CC ubiquitin acts as a degradation signal. Proteins which bind the  
 CC proteasome (a multisubunit complex) have ubiquitin-like (UBL)  
 CC domains. When the UBL domains are fused to a reporter protein,  
 CC they can be used to assess the proliferative potential and select  
 CC drug resistance of malignant cells. The UBL domain also allows the  
 CC purification of proteasomes for the isolation and characterisation

CC of novel subunits of the proteasome.  
XX Sequence 95 AA;  
SQ

Query Match 44.2%; Score 232; DB 20; Length 95;  
Best Local Similarity 45.4%; Pred. No. 5,1e-20;  
Matches 44; Conservative 21; Mismatches 28; Indels 4; Gaps 1;

QY 1 MSQDEKPPSTEDGDKQGEYIKLYIGDSSSEHFVKMTTLKTKESYCORQGVPMN 60  
D 1 madekpregyk-----temdhnlkvaqdgsvvqfklkrhplsklmkayceqglsmr 56  
QY 61 SLRFLFGORIAADNHPKELGMEEDVIEVYQRTG 97  
D 57 qirfrfdgplnetdtpqglemedtdidvfqgqtg 93

RESULT 7  
AAG08340 ID AAG08340 standard; Protein; 100 AA.  
AC AAG08340;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5836.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123588.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145319.  
PR 28-JUL-1999; 99US-0145351.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151086.  
 PR 30-AUG-1999; 99US-0151080.  
 PR 31-AUG-1999; 99US-0151303.  
 PR 01-SEP-1999; 99US-0151438.  
 PR 07-SEP-1999; 99US-0151930.  
 PR 10-SEP-1999; 99US-0152363.  
 PR 13-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 22-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160885.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161933.

PR 29-OCT-1999; 99US-0162142.  
 Query Match 42.6%; Score 223.5; DB 21; Length 100;  
 Best Local Similarity 51.2%; Pred. No. 5, 6e-19;  
 Matches 44; Conservative 14; Mismatches 25; Indels 3; Gaps 1;  
 QY 15 DKQGE---YIKAKVIGQDSSEIHFKVMTTHLKKKESTYCORQGVPMNSLRFLFEGQRI 71  
 DB 8 dkKpYdggaHlnKvkqgdqnevffIKrStqtklImnaycdYsvdmnsIatllfdgrtl 67  
 QY 72 ADNHTPKELGMEEDVLEVYQEOGTG 97  
 DB 68 raegtPdeldmedgdeldamlnhtg 93  
 RESULT 8  
 ID AAG1141 standard; Protein: 103 AA.  
 AC AAG1141;  
 XX 17-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 9747.  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 PN EP103405-A2.  
 XX 06-SEP-2000.  
 PD 25-FEB-2000; 2000EP-0301439.  
 PE 25-FEB-1999; 99US-0121825.  
 PF 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0128845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131440.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.



PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145917.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 20-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151303.  
PR 30-AUG-1999; 99US-0151438.  
PR 31-AUG-1999; 99US-0151930.  
PR 01-SEP-1999; 99US-0152363.  
PR 07-SEP-1999; 99US-0153070.  
PR 10-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154479.  
PR 20-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 29-SEP-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157753.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0161404.  
PR 22-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

42.1%; Score 221; DB 21; Length 103;



PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	15-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151068.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Oy		15 DKK--GGGYIKLAKYIGDSSIIHFKKMTTHLKKTKEGYCORGVPNMSIRFLFEGGRIA 72
Dd		8 dkkpdqgahnlkxkgggdgnevfifrikstcqlkximaycdrgsvdnslafldgrlr 67
Oy		73 DNHTPKELGMEEDVIEYYEOTGG 97
Dd		68 aeqptdelemedgdeldamlhtqts 92
RESULT	10	
ID	AAAG46643	
XX	AAAG46643 standard; Protein: 153 AA.	
AC	AAAG46643;	
DT	18-OCT-2000 (first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 58703.	
XX	Protein identification assay; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter	
KM	termination sequence.	
OS	Arabidopsis thaliana.	
XX	EPI033405-A2.	
PN	06-SEP-2000.	
PD	06-SEP-2000.	
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	15-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	
PR	23-APR-1999; 99US-0130510.	
PR	28-APR-1999; 99US-0130891.	
PR	30-APR-1999; 99US-0131449.	
PR	30-APR-1999; 99US-0132048.	
PR	04-MAY-1999; 99US-01332407.	
PR	05-MAY-1999; 99US-01332484.	
PR	06-MAY-1999; 99US-0133485.	
PR	06-MAY-1999; 99US-0133486.	
PR	07-MAY-1999; 99US-0132487.	
PR	11-MAY-1999; 99US-0132863.	
PR	14-MAY-1999; 99US-0134256.	
PR	14-MAY-1999; 99US-0134218.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134370.	
PR	18-MAY-1999; 99US-0134768.	
PR	19-MAY-1999; 99US-0134941.	
PR	20-MAY-1999; 99US-0135124.	
PR	21-MAY-1999; 99US-0135353.	
PR	24-MAY-1999; 99US-0135629.	
PR	27-MAY-1999; 99US-0136021.	
PR	27-MAY-1999; 99US-0136382.	
PR	28-MAY-1999; 99US-0136782.	
PR	01-JUN-1999; 99US-0137222.	
PR	03-JUN-1999; 99US-0137528.	
PR	04-JUN-1999; 99US-0137502.	
PR	07-JUN-1999; 99US-0137724.	
PR	08-JUN-1999; 99US-0138094.	
PR	10-JUN-1999; 99US-0138540.	
PR	10-JUN-1999; 99US-0138847.	

```
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139497.
PR 18-JUN-1999; 99US-0139497.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 28-JUL-1999; 99US-0145519.
PR 28-JUL-1999; 99US-0145519.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.1%; Score 221; DB 21; Length 153;
Best Local Similarity 51.8%; Pred. No. 1.9e-18;
Matches 44; Conservative 13; Mismatches 26; Indels 2; Gaps 1;
```



[illegible]



PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137722.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	14-JUN-1999;	99US-0138847.
PR	16-JUN-1999;	99US-0139119.
PR	17-JUN-1999;	99US-0139452.
PR	17-JUN-1999;	99US-0139453.
PR	18-JUN-1999;	99US-0139452.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139753.
PR	22-JUN-1999;	99US-0139817.
PR	23-JUN-1999;	99US-0139859.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	28-JUN-1999;	99US-0140695.
PR	29-JUN-1999;	99US-0140823.
PR	30-JUN-1999;	99US-0140991.
PR	01-JUL-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	02-JUL-1999;	99US-0142154.
PR	06-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	09-JUL-1999;	99US-0142803.
PR	12-JUL-1999;	99US-0142920.
PR	13-JUL-1999;	99US-0142977.
PR	14-JUL-1999;	99US-0143524.
PR	15-JUL-1999;	99US-0143624.
PR	16-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145226.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149375.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	25-AUG-1999;	99US-0149930.
PR	26-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159325.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	29-OCT-1999;	99US-0162142.

Query Match

33.4%; Score 175.5; DB 21; Length 111;





...

```
out_format : pfs
```

us-09-484-964-2.rge

OM of: US-09-40 5:57 PM  
Aug 15, 2001

Date: Aug 15, 2000  
Results were produced by CompuLink  
11835-26790/app\_query1  
11835-26790-000

About: Resonance  
Copyright (c)  
Parameters:  
-DEV=xlp  
-GAPEXT=4  
-runat=14082001-1100  
-POP=12.000  
-POP=4.500

[illegible]

Search US-09-484  
Query: length: 101  
emb1: \*

```

Query length: 10
Query length: 1344157
Database: GenEmbl: *
Database sequences: 856060004
Database length: 1074.650000
Database time (sec):
Search time (sec):

```

[illegible][illegible]

us-09-484-964-2.rge

[illegible]

alignment\_scores: 113 c  
Percent Similarity: 522.00  
alignment\_block: 5,168  
US-03-484-964-2 x HSSMT3C  
Align seg 1/21 to  
Percent Identity: 99.010  
Gaps: 101  
0  
13 others

1 MetSerASPGLNGLLALALySPGSGerTrGLuSPLeuGLyASPbLySLy 17  
42 ATGTCTGACCCAGAGAGCAAAACCTTCACACTGAGAGACTTGGGGGATAGGAA 91

seq\_documentation\_block:  
LOCUS HSU67122  
DEFINITION Human ubiquitin-related protein SUMO-1 mRNA, complete cds.  
ACCESSION HSU67122  
VERSION U67122  
KEYWORDS U67122.1 GI:1762972  
SOURCE human  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS 1 (bases) Eutheria; Chordata; Craniata; Vertebrata;  
TITLE Mahajan, R. I. to 816 A. Mahajan, R. I. to 816 A. Mahajan, R. I. to 816 A.  
JOURNAL Cell 68 (1), 97-107 (1997)  
MEDLINE 97148992  
REFERENCE 2 (bases 1 to 816) Mahajan, R. K., Guan, T., Delphin, C., Gerace, L. and Melchior, F.  
AUTHORS Direct S.R.K., Guan, T., Delphin, C., Gerace, L.  
JOURNAL Submitted (16-AUG-1996) Department of Molecular Biology, Research Institute  
FEATURES

```

source
Location/Qualifiers
1. 816
/organism="Homo sapiens"
/db_xref="taxon:9606"
81. 386 IMAGE Consortium
note="conjugated post-translationally to Rangapri;
Geoglutin-related protein;
Accession Number U03774; similar to UEL1 encoded
/codon_start=1
/product_start=1
protein="SUO-1"
/db_xref="id:AA050966.1"
translation="GI:1762973.1"
153 a KATSEYCOROGVPAKNSLRPLRSTEDGKKEGYIKATLYVIGDSSSEIKYKWKATYTH
173 g 235 t
alignmentscores:
Quality: 522.00
Ratio: 5.168

```

Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
US-09-484-964-2 x HS067122 ..

Align seg 1/1 to: HS067122 from: 1 to: 816

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysI 17
  |||
81 ATGCTCTCAAGCAGAGGCAAACTTCACATGAGGAGGATTAAGAA 130
  |||
17 sGlnGlyGluTyrTrlLeuLysLeuValIleGlyGlnAspSerSerGlu 34
  |||
131 GGAAGGTGCAATATTTAACTCAAGTCATTGGACAGAGATGAGAGAGA 180
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
  |||
181 TTCACCTCAAGTGAATGCAACACATCTCAAGAACTCAAGAAATCA 230
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPhe 67
  |||
231 TACTGTCAAGACAGAGGCTGTTCCAAATTCACATGAGGTTCTCTTGA 280
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
  |||
281 GGGTCAGAGAAATGCTGATATCATCTCCAAAGAACTGGGAATGAGG 330
  |||
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyHisSerThr 100
  |||
331 AAGAAGATGTGATTGATTCAGCAACAAACGGGGGTCATTCAACA 380
  |||
101 Val 101
  |||
381 GTT 383

```

seq\_name: gb\_pr10:HS038784

seq\_documentation\_block:

LOCUS HS038784 1017 bp mRNA PRI 16-OCT-1996

DEFINITION Human ubiquitin-like protein mRNA, complete cds.

ACCESSION U08784

VERSION U08784.1 GI:1574947

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1017)

AUTHORS Shen,Z., Pardington-Pourtymun,P.E., Comeaux,J.C., Moyzis,R.K. and Chen,D.J.

TITLE UBI1, a human ubiquitin-like protein associating with human

JOURNAL RAD51/RAD52 proteins

REFERENCE 2 (bases 1 to 1017)

AUTHORS Direct Submission

TITLE Submitted (16-OCT-1995) Zhiyuan Shen, Life Sciences Division, Los

JOURNAL Alamos National Lab, MS M888, Los Alamos, NM 87545, USA

FEATURES

source

1..1017

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/translation="MSDQKAPSTEDLGDKEGEYIKLVIGDSSSEIFKVKMTTHL  
KIKESYCGRGVPMNSLRFLFEGQRIADNHTKELGMEEDYIEVYQGTGSHSTV"

BASE COUNT 310 a 180 c 213 g 314 t

alignment\_scores:

quality: 522.00 Length: 101

Ratio: 5.168 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:

US-09-484-964-2 x HS038784 ..

Align seg 1/1 to: HS038784 from: 1 to: 1017

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysI 17
  |||
67 ATGCTCTCAAGCAGAGGCAAACTTCACATGAGGAGGATTAAGAA 116
  |||
17 sGlnGlyGluTyrTrlLeuLysLeuValIleGlyGlnAspSerSerGlu 34
  |||
117 GGAAGGTGCAATATTTAACTCAAGTCATTGGACAGAGATGAGAGAGA 166
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
  |||
167 TTCACCTCAAGTGAATGCAACACATCTCAAGAACTCAAGAAATCA 216
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPhe 67
  |||
217 TACTGTCAAGACAGAGGCTGTTCCAAATTCACATGAGGTTCTCTTGA 266
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
  |||
267 GGGTCAGAGAAATGCTGATATCATCTCCAAAGAACTGGGAATGAGG 316
  |||
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyHisSerThr 100
  |||
317 AAGAAGATGTGATTGATTCAGCAACAAACGGGGGTCATTCAACA 366
  |||
101 Val 101
  |||
367 GTT 369

```

seq\_name: gb\_r01:AF033353

seq\_documentation\_block:

LOCUS AF033353 1187 bp mRNA ROD 06-FEB-1998

DEFINITION Mus musculus ubiquitin-homology domain protein (Ubl1) mRNA,

complete cds.

ACCESSION AF033353

VERSION AF033353.1 GI:2645736

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1187)

AUTHORS Howe,K., Williamson,J., Boddy,N., Sheer,D., Freemont,P. and Solomon,E.

TITLE The ubiquitin-homology gene P1C1: characterization of mouse (P1C1)

JOURNAL JOURNAL

REFERENCE 2 (bases 1 to 1187)

AUTHORS Howe,K., Freemont,P.S. and Solomon,E.

TITLE Direct Submission

JOURNAL Submitted (07-NOV-1997) Medical & Molecular Genetics, Guy's

/translation="MSDQKAPSTEDLGDKEGEYIKLVIGDSSSEIFKVKMTTHL  
KIKESYCGRGVPMNSLRFLFEGQRIADNHTKELGMEEDYIEVYQGTGSHSTV"

BASE COUNT 310 a 180 c 213 g 314 t

alignment\_scores:

quality: 522.00 Length: 101

Ratio: 5.168 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:

US-09-484-964-2 x HS038784 ..

Align seg 1/1 to: HS038784 from: 1 to: 1017

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysI 17
  |||
67 ATGCTCTCAAGCAGAGGCAAACTTCACATGAGGAGGATTAAGAA 116
  |||
17 sGlnGlyGluTyrTrlLeuLysLeuValIleGlyGlnAspSerSerGlu 34
  |||
117 GGAAGGTGCAATATTTAACTCAAGTCATTGGACAGAGATGAGAGAGA 166
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
  |||
167 TTCACCTCAAGTGAATGCAACACATCTCAAGAACTCAAGAAATCA 216
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPhe 67
  |||
217 TACTGTCAAGACAGAGGCTGTTCCAAATTCACATGAGGTTCTCTTGA 266
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
  |||
267 GGGTCAGAGAAATGCTGATATCATCTCCAAAGAACTGGGAATGAGG 316
  |||
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyHisSerThr 100
  |||
317 AAGAAGATGTGATTGATTCAGCAACAAACGGGGGTCATTCAACA 366
  |||
101 Val 101
  |||
367 GTT 369

```

seq\_name: gb\_r01:AF033353

seq\_documentation\_block:

LOCUS AF033353 1187 bp mRNA ROD 06-FEB-1998

DEFINITION Mus musculus ubiquitin-homology domain protein (Ubl1) mRNA,

complete cds.

ACCESSION AF033353

VERSION AF033353.1 GI:2645736

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1187)

AUTHORS Howe,K., Williamson,J., Boddy,N., Sheer,D., Freemont,P. and Solomon,E.

TITLE The ubiquitin-homology gene P1C1: characterization of mouse (P1C1)

JOURNAL JOURNAL

REFERENCE 2 (bases 1 to 1187)

AUTHORS Howe,K., Freemont,P.S. and Solomon,E.

TITLE Direct Submission

JOURNAL Submitted (07-NOV-1997) Medical & Molecular Genetics, Guy's

```

/db_xref="taxon:10090"
/chr="2-3"
1..1187
/gene="Ubl1"
/ncbi="P1"
119..424
/gene="Ubl1"
/codon_start=1
/product="ubiquitin-homology domain protein"
/protein_id="AAC39959.1"
/db_xref="GI:2645737"
/translation="MSDQKAPSTEDGDKKEGYIKLVIGDSSSEIHKVKMTTHL
KKIKESYCOQGVPMNSRLFLFGORADNHTPELGMEDVIEVYQEOGTGSHSTV"
BASE COUNT      350 a      216 c      267 g      354 t
ORIGIN
alignment_scores:
  Quality: 522.00      Length: 101
  Ratio: 5.168      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.010
alignment_block:
US-09-484-964-2 x AF033353 ..
Align seg 1/1 to: AF033353 from: 1 to: 1187
1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
119 ATGTCTGACGAGGAGGAGCAAAACCTTCACAGGAGGAGGAGGAGGAGGAG 168
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
169 GGAAGGAGAAATACATTAACCTCAAGTATTGAGAGGATGAGAGTGA 218
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
219 TACATTTCAAAAGTGAATAATGACACACATCTCAAGAAACCTCAAGAAATCA 268
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheGlu 67
269 TACTGTCAAAAGCAGGAGGAGTCCATTAATCACTCAGGTTTCTCTTGA 318
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
319 AGGTCACAGAAATGCTGATATCATCTCCGAAAGAACTGGCAATGAGG 368
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThr 100
369 AAGAAGATGTGATTGAAGTTATCAGAACAAACGGGGGCTCAGTCAAGC 418
101 Val 101
111
419 GTT 421
seq_name: gb_pr10:HSU61397
seq_documentation_block:
LOCUS      HSU61397      1223 bp      mRNA      PRI      14-JAN-1997
DEFINITION Human ubiquitin-homology domain protein P101 mRNA, complete cds.
ACCESSION  U61397
VERSION     U61397.1  GI:1518693
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1223)
AUTHORS     Boddy,M.N., Howe,K., Etkin,L.D., Solomon,E. and Freemont,P.S.
TITLE       P101, a novel ubiquitin-like protein which interacts with the PM1
component of a multiprotein complex that is disrupted in acute
promyelocytic leukaemia

```

```

JOURNAL      Oncogene 13 (5), 971-982 (1996)
MEDLINE      96400311
REFERENCE    2 (bases 1 to 1223)
AUTHORS     Howe,K., Boddy,M.N., Etkin,L.D., Solomon,E. and Freemont,P.S.
TITLE       Direct Submission
JOURNAL      Submitted (20-JUN-1996) Somatic Cell Genetics, Imperial Cancer
Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
FEATURES
  source
1..1223
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="Placenta"
146..451
/codon_start=1
/product="ubiquitin-homology domain protein P101"
/protein_id="BAB40388.1"
/db_xref="GI:1518694"
/translation="MSDQKAPSTEDGDKKEGYIKLVIGDSSSEIHKVKMTTHL
KKIKESYCOQGVPMNSRLFLFGORADNHTPELGMEDVIEVYQEOGTGSHSTV"
BASE COUNT      371 a      209 c      268 g      375 t
ORIGIN
alignment_scores:
  Quality: 522.00      Length: 101
  Ratio: 5.168      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.010
alignment_block:
US-09-484-964-2 x HSU61397 ..
Align seg 1/1 to: HSU61397 from: 1 to: 1223
1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
146 ATGTCTGACGAGGAGGAGCAAAACCTTCACAGGAGGAGGAGGAGGAGGAG 195
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
196 GGAAGGAGAAATACATTAACCTCAAGTATTGAGAGGATGAGAGTGA 245
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
246 TACATTTCAAAAGTGAATAATGACACACATCTCAAGAAACCTCAAGAAATCA 295
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheGlu 67
296 TACTGTCAAAAGCAGGAGGAGTCCATTAATCACTCAGGTTTCTCTTGA 345
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
346 GGGTCACAGAAATGCTGATATCATCTCCGAAAGAACTGGCAATGAGG 395
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThr 100
396 AAGAAGATGTGATTGAAGTTATCAGAACAAACGGGGGCTCAGTCAACA 445
101 Val 101
111
446 GTT 448
seq_name: gb_pr7:BC006462
seq_documentation_block:
LOCUS      BC006462      1227 bp      mRNA      PRI      13-APR-2001
DEFINITION Homo sapiens, ubiquitin-like 1 (sentrin), clone MGC:2095, mRNA,
complete cds.
ACCESSION  BC006462
VERSION     BC006462.1  GI:13623670
KEYWORDS    MGC.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Mammalia: Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 1227)

Strausberg, R.  
Direct Submission

Submitted (09-APR-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcsc.bc.ca](mailto:info@bcsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: ITAL Plate: 8 Row: f Column: 12.  
Location/Qualifiers

## FEATURES

Source

1. 1227  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:2095 IMAGE:3546125"  
/tissue.type="Placenta, Chorioncarcinoma"  
/clone.lib="NIH\_MGC\_21"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
132. 437  
/codon\_start=1  
/product="ubiquitin-like 1 (sentrin)"  
/protein\_id="AA06462.1"  
/db\_xref="GI:13623671"  
/translation="MSDQKRPSTEDLDKREGEYIKLVIGDSSSEIHFKVMTTHL  
KKLESYCORGVPMNSLRFLFEGRIADNHTPRELGEEDVLEVOEOTGHSYV"

CDS

BASE COUNT 386 a 208 c 259 g 374 t  
ORIGIN

## alignment\_scores:

Quality: 522.00 Length: 101  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.010

## alignment\_block:

US-09-484-964-2 x BC006462 ..

Align seg 1/1 to: BC006462 from: 1 to: 1227

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17  
|||||  
132 ATGCTGACCAAGGAGGCAAACTTCAACTGAGACTTGGGGATTAAGAA 181  
17 sGlnGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34  
|||||  
182 GGAAGGTGAATATTATTAAGTCAAGCATTTGGACAGATGACAGTGA 231  
34 LeuIlePheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50  
|||||  
232 TTCCTTCAAGTGAAGTGAACACATCTCAAGAACTCAAGATCA 281

51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67  
|||||  
282 TACTGTCAAGACAGAGGTGTTCCAAATGAAATTCACCTGAGCTTCTTTGA 331  
67 uGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84  
|||||  
332 GGGTCAGAAATTCCTGATATCATCTACATCAAAAGAACTGGGAATGAGG 381  
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThr 100  
|||||  
382 AAGAAGATGTGATGAGTTATTCAGAAACAAACGGGGGTTCATTCAACA 431  
101 Val 101  
432 GTT 434  
seq\_name: gb.pat1:AR087051

## seq\_documentation\_block:

LOCUS AR087051 1514 bp DNA PAT 07-SEP-2000  
DEFINITION Sequence 1 from patent US 5985664.  
ACCESSION AR087051  
VERSION AR087051.1 GI:10013817  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 1514)  
AUTHORS Baker, B.F. and Cowse, L.M.  
TITLE Antisense modulation of Sentrin expression  
JOURNAL Patent: US 5985664-A 1 16-NOV-1999;  
FEATURES  
Source  
1. 1514  
/organism="unknown"

BASE COUNT 462 a 256 c 314 g 482 t  
ORIGIN

## alignment\_scores:

Quality: 522.00 Length: 101  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.010

## alignment\_block:

US-09-484-964-2 x AR087051 ..

Align seg 1/1 to: AR087051 from: 1 to: 1514

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17  
|||||  
136 ATGCTGACCAAGGAGGCAAACTTCAACTGAGACTTGGGGATTAAGAA 185  
17 sGlnGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34  
|||||  
186 GGAAGGTGAATATTATTAAGTCAAGTCAATTCAGACAGATGACATGAGA 235  
34 LeuIlePheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50  
|||||  
236 TTCCTTCAAGTGAAGTGAACATGACACACATCTCAAGAACTCAAGATCA 285  
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67  
|||||  
286 TACTGTCAAGACAGAGGTGTTCCAAATGAAATTCACCTGAGCTTCTTTGA 335  
67 uGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84  
|||||  
336 GGGTCAGAAATTCCTGATATCATCTACATCAAAAGAACTGGGAATGAGG 385  
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThr 100  
|||||  
386 AAGAAGATGTGATGAGTTATTCAGAAACAAACGGGGGTTCATTCAACA 435  
101 Val 101

```

|||||
436 GTT 438

seq_name: gb_pri0:HSU83117

seq_documentation_block:
LOCUS      HSU83117      1514 bp      mRNA      PRI      09-JAN-1997
DEFINITION Human sentrin mRNA, complete cds.
ACCESSION  U83117
VERSION    U83117.1  GI:1769601
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1514)
            Okura,T., Gong,L., Kamitani,T., Wada,T., Okura,I., Wei,C.-F.,
            Chang,H.-M. and Yeh,E.T.H.
            Protection Against Fas/Apo-1- and Tumor Necrosis Factor-Mediated
            Cell Death by a Novel Protein, Sentrin
JOURNAL    J. Immunol. 157(10), 4277-4281 (1996)
AUTHORS    2 (bases 1 to 1514)
            Okura,T., Gong,L., Kamitani,T., Wada,T., Okura,I., Wei,C.-F.,
            Chang,H.-M. and Yeh,E.T.H.
            Direct Submission
TITLE       Submitted (23-DEC-1996) Division of Molecular Medicine, Department
            of Internal Medicine, and Cardiovascular Research Center, Institute
            of Molecular Medicine for the Prevention of Human Diseases, The
            University of Texas-Houston Health Science Center, 2121 W. Holcombe
            Blvd., Suite 900, Houston, TX 77030, USA
FEATURES   Location/Qualifiers
            source          1..1514
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /cissue_type="placenta"
                           136..441
                           /note="ubiquitin-like protein"
                           /codon_start=1
                           /product="sentrin"
                           /protein_id="AA839698.1"
                           /db_xref="GI:1769602"
                           /translation="MSDOEAKPSTEDLGKKEGEYIKLVIGDSSSEIHFKVMTHTL
                           KKLKESYCORGVPMNSLRFLFEGORIADNHTPELGMEEEDVLEVOEQTGSHSTV"
BASE COUNT 462 a      256 c      314 g      482 t
ORIGIN
alignment_scores:
Quality: 522.00      Length: 101
Ratio: 5.168          Gaps: 0
Percent Similarity: 100.000      Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x HSU83117 ..

Align seg 1/1 to: HSU83117 from: 1 to: 1514

1 MetsetASpGLngLualalysProserThrguAspLeuGLyAspLysly 17
|||||
136 ATGTGTGACGACGAGGCAAAACCTTCACCTGAGACTGGGGGATAGAA 185
17 scLngLyuTyrlleYsleuYsValleGlyGlnAspSerSerGluI 34
|:::|||||
186 GGAAGTGAAATATTAACCTCAAGTCATGTGACAGATAGAGTGA 235
34 lshisPhelysVallysWecthrThHisLeuYsLysleuYsgLuser 50
|||||
236 TTCACCTCAAAAGTGAAGAAATGACACATCTCAAGAACTCAAGAAATCA 285
51 TTYCYSGlnArGlnGlyAlaPrometAsnSerLeuArGPhelenuheG 67
|||||
286 TACTGTCAAGACAGGGGTGTTCCAAATGAATCACTCAGGTTTCTCTTGA 335

```

```

67 ucLglnArGylleAlaAspAsnHisThrProLysGluLeuGlyMeTcLug 84
|||||
336 GGCTCAGAGATTGCTGATNATCATCTCCAAAGAACTGGGAATGAGG 385
84 luGluAspValleGlyValTyrlGlnGluGlnThrgLysHisSerThr 100
|||||
386 AAGACAGTGTGATTGAAGTTATTCAGAGAACAAACGGGGGTCTATTCACA 435
101 Val 101
|||
436 GTT 438

seq_name: gb_htg17:AC084066

seq_documentation_block:
LOCUS      AC084066      235411 bp      DNA      HTG      12-OCT-2000
DEFINITION Mus musculus clone RP23-321D1, *** SEQUENCING IN PROGRESS ***, 29
            unordered pieces.
ACCESSION  AC084066
VERSION    AC084066.1  GI:10799415
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 235411)
            DOE Joint Genome Institute.
            Sequencing of Mouse
            Unpublished
JOURNAL    2 (bases 1 to 235411)
            DOE Joint Genome Institute.
AUTHORS    Direct Submission
TITLE       Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 2351294
            Center clone name: RPCI-23_321D1
            -----
            Summary Statistics
            Consensus quality: 214207 bases at least Q40
            Consensus quality: 223053 bases at least Q30
            Consensus quality: 225208 bases at least Q20
            Estimated insert size: 200000; pulse field gel estimation
            Estimated insert size: 232611; sum-of-contigs estimation
            Quality coverage: 11.56 in Q20 bases; pulse field gel estimation
            Quality coverage: 9.94 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *
            1 1210: contig of 1210 bp in length
            * 1211
            * 1310: gap of unknown length
            * 1311
            * 3240: contig of 1930 bp in length
            * 3241
            * 3340: gap of unknown length
            * 3341
            * 4504: contig of 1164 bp in length
            * 4505
            * 4604: gap of unknown length
            * 4605
            * 6172: contig of 1568 bp in length
            * 6173
            * 6272: gap of unknown length
            * 6273
            * 7354: contig of 1082 bp in length
            * 7355
            * 7454: gap of unknown length
            * 7455
            * 8625: contig of 1171 bp in length
            * 8626
            * 8725: gap of unknown length
            * 8726
            * 10114: contig of 1389 bp in length
            * 10214: gap of unknown length

```



```

113006 TACATTTCAAAATGAAATGACACACATCTCACAACAACTCCAAAATCA 112957
51 TTYCYSGINATgInGInGlyValPromeKtsnSerLseuAqPhLeuPhegl 67
|||||
112956 TACTCTCAAGACAGGAGGATTCCAATGAATTCACACAGGATTCCTTTGA 112907
67 uG1gLnATrg11lea1aSPaNHs1ThProlYsg1luenuglyMetgUG 84
|||||
112906 AGGTCAGAGAAATTCGTGTATATCATACATCCGAAAGAACTGGGAAATGAGA 112857
84 lUG1uASPvAl11eGluValTyrgInGInGInThrGlyGlyH1SerThr 100
|||||
112856 AAGAAGATGTGAATGAAGTTATCAGAAACAAGGGGGTCACTCGACG 112807
101 Val 101
112806 GTT 112804
seq_name: gb_om:AF242526
seq_documentation_block:
LOCUS AF242526 510 bp mRNA MAM 03-NOV-2000
DEFINITION Cervus nippon sentrin mRNA, complete cds.
ACCESSION AF242526
VERSION AF242526.1 GI:9664276
KEYWORDS
SOURCE
ORGANISM
sika deer.
Cervus nippon
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;
Cervidae; Cervinae; Cervus.
1 (bases 1 to 510)
Sun,L.G., Yu,Y.L. and Jiang,Y.
Direct Submission
Submitted (07-MAR-2000) Dept. of Immunology, Norman Bethune
University of Medical Sciences, Xin Min Street, Changchun, Jilin
130021, People's Republic of China
location/Qualifiers
1..510
/organism="Cervus nippon"
/db_xref="taxon:9863"
/note="authority: Cervus nippon Temminck"
129..434
/codon_start=1
/product="sentrin"
/protein_id="AA97049.1"
/db_xref="GI:9664277"
/translation="MSDQEAKPSTEDLDGKKEGEYIKLVIGDSSLEHFKWMTTHL
KKLKESTYCORGVYMNSLRFLFEQORJADNHTPRELMEEEDVIEYVHEDTGSHIV"
BASE COUNT 154 a 98 c 121 g 137 t
ORIGIN

```

```

alignment_scores:
    Quality: 517.00
    Ratio: 5.119
    Gaps: 0
    Percent similarity: 100.000
    Percent identity: 98.020
alignment_block:
    US-09-484-964-2 x AF242526 ..
Align seg 1/1 to: AF242526 from: 1 to: 510
1 MetSerAspGInGluAlaIalysProSerThrGluAspLeuGlyAspLysly 17
|||||
129 ATGTCGTGACCAAGAAACCTTCAACCGAGACTTGGGGGATTAAGAA 178
|||||
17 sGInGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
179 GGAAGGGAATATATTAACCTCAAAAGTCATTGCACAGAGATACGATAGA 228
|||||
34 LeHISpEluLysValIyMetThrThrIstLeuLysLysLeuLysGluSer 50
|||||

```

```

229 TTCACCTCAAGATGAAATGACGACACATCTCAAGAACTCAAGAAATCA 278
51 TTYCYSGlnArgGlnGlnValPrometAsnSerLeuArgPheLeuPheG1 67
|||||
279 TACTGTCAAGACAGGAGGTCTTATGAAATTCACAGGTTCTCTTTGA 328
|||||
67 uGlyGlnAArgIleAlaAspAsnHisThrProlySGluLeuGlyMetGluG 84
|||||
329 AGGTCAAGAAATGCTGATATCACTCAAAAGAACTGGGAATGAGG 378
|||||
84 luGluAspValIleGluValTyrGlnGlnGlnThrGlyGlnHisSerThr 100
|||||
379 AAGAAATGATTGATTGAACTTATCATGACAAACAGGGGTCATTCACG 428
|||||
101 Val 101
429 GTT 431

seq_name: gb_hgt22:AL513282

seq_documentation_block:
LOCUS AL513282 155913 bp DNA HTG 26-FEB-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-179G5, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION AL513282
VERSION AL513282.5 GI:13162044
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 155913)
Pavitt,R.
Direct Submission
Submitted (25-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clone requests@sanger.ac.uk
On Feb 28, 2001 this sequence version replaced gi:13161719.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: DA17965
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 153015 bases at least Q40
Consensus quality: 154565 bases at least Q20
Insert size: 155213; sum-of-contrigs
Insert size: 164818; 1.0% error; agarose-1p
Quality coverage: 6.89x in Q20 bases; sum-of-contrigs quality
coverage: 7.27x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 20838: contrig of 20838 bp in length
* 20939: gap of 100 bp
* 50589: contrig of 29651 bp in length
* 50590: gap of 100 bp
* 50690: 50689; gap of 100 bp
* 81538: contrig of 30849 bp in length
* 81539: 81638; gap of 100 bp
* 81639: 89682; contrig of 8044 bp in length
* 89683: 89782; gap of 100 bp

```

```

FEATURES
Source
1..155913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-179G5"
/clone_1fb="RP11-1.1"
1..20838
/note="assembly_fragment:02743
fragment_chain:1
clone_end:17
vector_side:left"
20939..50589
/note="assembly_fragment:01236
fragment_chain:1"
50690..81538
/note="assembly_fragment:00478
fragment_chain:2"
81639..89682
/note="assembly_fragment:03283
fragment_chain:2"
89783..108105
/note="assembly_fragment:00942
fragment_chain:3"
108206..149395
/note="assembly_fragment:02750
fragment_chain:3"
149496..153075
/note="assembly_fragment:02486
fragment_chain:4"
153176..155913
/note="assembly_fragment:03083
fragment_chain:4"
BASE COUNT 43830 a 34529 c 35326 g 41527 t 701 others
ORIGIN
alignment_scores:
Quality: 512.00 Length: 101
Ratio: 5.120 Gaps: 0
Percent Similarity: 99.010 Percent Identity: 97.030
alignment_block:
US-09-484-964-2 x AL513282/rev ..
Align seg 1/1 to reverse of: AL513282 from: 1 to: 155913
1 MetSerAspGlnGlnAlaLysProSerThrGluAspLeuGlyAspLysly 17
|||||
136335 ATGCTGACCAAGGAGGCAAACTTCACTGAGGACTTGGGGATGAAGA 136306
|||||
17 sGlnGlyLysTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
136305 GCAAGGTGATATATTAATCACTCAATGACAGATGACAGTACGAGA 136256
|||||
34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
136255 TTCACCTCAAGATGAAATGACACACATCTCAAGAACTCAAGAAATCA 136206
|||||
51 TTYCYSGlnArgGlnGlnValPrometAsnSerLeuArgPheLeuPheG1 67
|||||
136205 TACTGTCAAGACAGGCGCTTCAATGAAATTCATCAGGTTCTCTTTGA 136156
|||||
67 uGlyGlnAArgIleAlaAspAsnHisThrProlySGluLeuGlyMetGluG 84
|||||
136155 GGGTCAAGAAATGCTGATATCATCTCAAAAGAACTGGGAATGAGG 136106

```

KEY WORDS .

KEY WORDS .

SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 372)  
AUTHORS O'Neill, R. and Palase, P.  
TITLE Antiviral compounds that inhibit interaction of host cell proteins and viral proteins required for replication  
JOURNAL Patent: EP 0861322-A 18 02-SEP-1998;  
MOUNT SINAI MEDICAL CENTER (US)  
FEATURES Location/Qualifiers  
1..372  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 121 a 72 c 97 g 82 t  
ORIGIN

alignment\_scores:  
Quality: 492.00 Length: 97  
Ratio: 5.125 Gaps: 0  
Percent Similarity: 98.969 Percent Identity: 97.938

alignment\_block:  
US-09-484-964-2 x AX021844 ..  
Align seg 1/1 to: AX021844 from: 1 to: 372

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
80 ATGCTGACCCAGGAGCGCAAACTTCAACTGAGACTGGGGGATTAAGAA 129
17 sGlnGlyLysValIleuLysLeuLysValIleGlyGlnAspSerSergLys 34
130 GGAAGGCAATATATTAACTCAAAAGTCATTGGACAGATAGCAGTGAAGA 179
34 LeuIleHisPheLysValLysMetThrThrHisLysLysLysLysLysLys 50
180 TTCCTCTCAAAAGTCGAAATGACACACATCTCAAGAACTCAAGAAATCA 229
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheGlu 67
230 TACTGTCAAGACAGGCTGTTCCATGAAATTCATCAGGTTCTCTTGA 279
67 uGlnArgLysIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
280 GGGTCAGAGAAATGCTGATATCATCTCCAAAGAACTGGGAATGGAGG 329
84 LuGluAspValIleGluValTyrGlnGlnGlnGlnThrGlyGly 97
330 AAGAAGCTCTGATGAGTTTATCAGGAACAAGGGGGGT 370

```

seq\_name: gb\_ov: XLSUMO

seq\_documentation\_block:  
LOCUS XLSUMO 466 bp mRNA VRT 16-JAN-1998  
DEFINITION Xenopus laevis SUMO-1 mRNA, partial cds.  
ACCESSION Z97073  
VERSION Z97073.1 GI:2791891  
KEYWORDS SUMO-1 gene.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 466)  
Saitoh, H., Sparrow, D.B., Shiomi, T., Pu, R.T., Nishimoto, T.,  
Mohn, T.J. and Dasso, M.  
TITLE Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2  
JOURNAL Curr. Biol. 8 (2), 121-124 (1998)  
MEDLINE 98089176  
REFERENCE 2 (bases 1 to 466)  
Sparrow, D.B.  
TITLE Direct Submission

JOURNAL Submitted (19-JUN-1997) Developmental Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA,  
United Kingdom  
FEATURES Location/Qualifiers  
source 1..466  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
gene 58..366  
/gene="SUMO-1"  
CDS 58..366  
/gene="SUMO-1"  
/codon\_start=1  
/product="SUMO-1 protein"  
/protein\_id="CAB09807.1"  
/db\_xref="GI:2791892"  
/db\_xref="SPTREMBL:O57686"  
/translation="MSDQEARPSSEDLGDKKGGDYIKLVIGDSSIEHFVKMTHT  
LKKLESYROROGVPMNSIRLFLPEGORISDHPRELGMEDVDVLEVYEOETGGHSTF"

BASE COUNT 138 a 89 c 102 g 137 t  
ORIGIN

alignment\_scores:  
Quality: 470.50 Length: 101  
Ratio: 4.753 Gaps: 1  
Percent Similarity: 98.020 Percent Identity: 92.079

alignment\_block:  
US-09-484-964-2 x XLSUMO ..  
Align seg 1/1 to: XLSUMO from: 1 to: 466

```

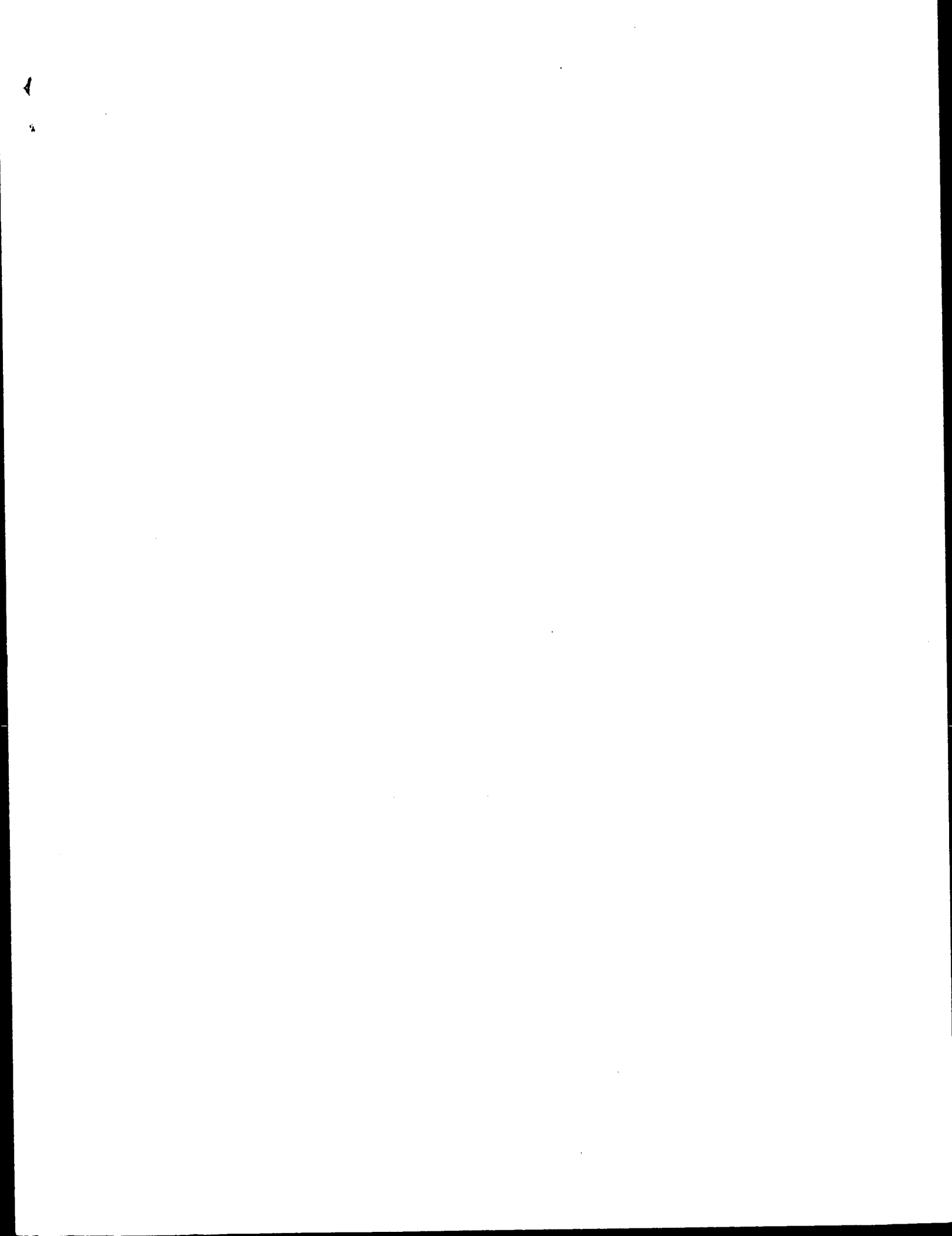
1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
58 ATGCTGATCAGGAAGCAAAACCATCTAGCGAGATCTAGGAGACAAAAA 107
17 sGln...GlyLysValIleuLysLeuLysValIleGlyGlnAspSerSerg 33
108 AGATGAGAGGGATATATCAAACTCAAGTCATTGGACAGGACAGAGTG 157
33 LuIleHisPheLysValLysMetThrThrHisLysLysLysLysLysLys 49
158 AAATTCATCTCAAGGTAAGATGACACACCATCTTAAAAAGCTGAAGAG 207
50 SerTyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPhe 66
208 TCATACCGTCAGAGACAGCGCGTCCATGAATTCCTCAGGTTTGT 257
66 eGlnGlyGlnArgLysIleAlaAspAsnHisThrProLysGluLeuGlyMetG 83
258 TGAAGGCAAGAGATCTCAGATCACAGACTCTTAAGAGAGCTGGAAATGG 307
83 LuGluLysAspValIleGluValTyrGlnGlnGlnGlnThrGlyGlyHisSer 99
308 AGCAAGAGATGTTATTGAAGTTATCAGGACAGAGCTGGGGCCACTCG 357
100 Thr 100
|||
358 ACA 360

```

Thu Aug 16 08:48:53 2001

us-09-484-964-2.rge

---





```
|||||
88 ATGTCGACCGAGGCAAAACCTTCACTGAGACTTGCGGATGAAGAA 137
17 sGlnGlyLeuTyrIleLeuLeuValIleGlyLysSerSerGlu 34
138 GCAGAGTGATATATTAACTCAAGTCATTGACAGATAGCACTGAGA 187
34 LeHISpHeLysValLysMetThrThrHISLeuLysLeuGlyLysSer 50
188 TTCACCTCAAACTGAAATGACACATCTCCAGAACTCAAGAAATCA 237
51 TYTCysGlnArgGlnGlyValProMetLysSerLeuArgPheLeuPheG 67
238 TACTGTCAAAAGCAGAGGTGTTCCAAATGAAATCAGCTGCTCTTGA 287
67 uGlyLysArgIleAlaAspAsnHISThrProLysGluLeuGlyMetLug 84
288 GGGTCAGAGAAATGCTGTATATCACTACTCCAAAGAACTGGAAATGAG 337
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHISSerThr 100
338 AAGAAGATGATGATTGAAGTTTATCGAGCAACAAAGCGGCTCATTCACA 387
101 Val 101
388 GTT 390

seq_name: /SIDSJ/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV59722
seq_documentation_block:
ID AAV59722 standard; DNA; 1196 BP.
XX
AC AAV59722;
XX
DT 19-JAN-1999 (first entry)
XX
DE Human secreted protein gene 47 clone HOGAV75.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9839448-A2.
XX
PD 11-SEP-1998.
XX
PE 06-MAR-1998; 98WO-US04493.
XX
PR 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
```

```
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 23-MAY-1997; 97US-0047634.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
```



PR 22-AUG-1997; 97US-0056908.  
 PR 22-AUG-1997; 97US-0056909.  
 PR 22-AUG-1997; 97US-0056910.  
 PR 22-AUG-1997; 97US-0056911.  
 PR 05-SEP-1997; 97US-0057650.  
 PR 05-SEP-1997; 97US-0057659.  
 PR 05-SEP-1997; 97US-0057761.  
 PR 12-SEP-1997; 97US-0058785.  
 XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;  
 PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;  
 PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 XX

DR MPI, 1998-506364/43.  
 DR P-PDB; AAM74938.

XX New isolated human genes and the secreted polypeptide(s) they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders

PS Claim 1: Page 449; 721pp; English.

XX This sequence represents a nucleic acid molecule designated Gene 47 from  
 CC the human cDNA clone HOGAV75 (deposited as clone ATCC 97897 and ATCC  
 CC 209043) which encodes a secreted human protein. The gene can be used to  
 CC generate fusion proteins by linking to the gene to a human  
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of  
 CC the fused protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: AAV59511-V59812; amino acid sequences AAM74731-W5026)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological  
 CC conditions can be diagnosed by determining the amount of the new  
 CC polypeptides in a sample or by determining the presence of mutations in  
 CC the new polynucleotides. Specific uses are described for each of the 186  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAV59511 for described uses).  
 XX

SQ Sequence 1196 BP; 360 A; 212 C; 259 G; 364 T; 1 other;

alignment\_scores:

Quality: 522.00 Length: 101  
 Ratio: 5.168 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:

US-09-484-964-2 x AAV59722 ..

Align seg 1/1 to: AAV59722 from: 1 to: 1196

1 MetSerAspGlnGlnAlaLysProSerThrGluAspLeuGlyAspLysIle 17  
 134 ATGCTGACACAGAGCAAAACCTTCACAGAGCTGGGGATTAAGAA 183  
 17 sgInglyGluTyrIleLeuLysValIleGlyGlnAspSerSerGluI 34  
 184 GGAAGCGGAATATATTAACTCAAGTCATTGGACAGGATGACAGGAGA 233  
 34 LeuHisPheValLysMetThrThrHisLeuLysLysLeuLysGluSer 50  
 234 TTCACCTCAAGTGAATATCAACACATCTCAAGAAATCAAGAAATCA 283  
 51 TyrCysGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheG 67  
 284 TACTGCTAAAGACAGGGGTTCATCAATGATTCACAGCTTCTCTTGA 333  
 67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84  
 334 GGGTCAGAGAAATTCGATATATCACTCCAAAGAACTGGGAATGAGAG 383

84 LuLuspyValIleGlyValTyrGlnGlnGlnThrGlyLysIleSerThr 100  
 384 AAGAGATGATTTGAAGTTTATTCAGGAACAAACGGGGGTTCATTCACA 433  
 101 Val 101  
 434 GTT 436

seq\_name: /SID51/gcdata/geneseq/geneseqn/NA1998.DAT:AAV59557

seq\_documentation\_block:

ID AAV59557 standard; DNA; 1220 BP.

AC AAV59557;

DE 06-JAN-1999 (first entry)

XX Human secreted protein gene 47 clone HOGAV75.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

PN W09839448-A2.

XX 11-SEP-1998.

PF 06-MAR-1998; 98WO-US04493.

PR 02-OCT-1997; 97US-0061060.  
 PR 07-MAR-1997; 97US-0038621.  
 PR 07-MAR-1997; 97US-0040161.  
 PR 07-MAR-1997; 97US-0040162.  
 PR 07-MAR-1997; 97US-0040163.  
 PR 07-MAR-1997; 97US-0040333.  
 PR 07-MAR-1997; 97US-0040334.  
 PR 07-MAR-1997; 97US-0040336.  
 PR 07-MAR-1997; 97US-0040626.  
 PR 11-APR-1997; 97US-0043311.  
 PR 11-APR-1997; 97US-0043312.  
 PR 11-APR-1997; 97US-0043313.  
 PR 11-APR-1997; 97US-0043314.  
 PR 11-APR-1997; 97US-0043568.  
 PR 11-APR-1997; 97US-0043569.  
 PR 11-APR-1997; 97US-0043576.  
 PR 11-APR-1997; 97US-0043578.  
 PR 11-APR-1997; 97US-0043659.  
 PR 11-APR-1997; 97US-0043669.  
 PR 11-APR-1997; 97US-0043670.  
 PR 11-APR-1997; 97US-0043671.  
 PR 11-APR-1997; 97US-0043672.  
 PR 11-APR-1997; 97US-0043674.  
 PR 23-MAY-1997; 97US-0047492.  
 PR 23-MAY-1997; 97US-0047500.  
 PR 23-MAY-1997; 97US-0047501.  
 PR 23-MAY-1997; 97US-0047502.  
 PR 23-MAY-1997; 97US-0047503.  
 PR 23-MAY-1997; 97US-0047581.  
 PR 23-MAY-1997; 97US-0047582.  
 PR 23-MAY-1997; 97US-0047583.  
 PR 23-MAY-1997; 97US-0047584.  
 PR 23-MAY-1997; 97US-0047585.  
 PR 23-MAY-1997; 97US-0047586.  
 PR 23-MAY-1997; 97US-0047587.  
 PR 23-MAY-1997; 97US-0047588.

PR 23-MAY-1997; 97US-0047589.  
 PR 23-MAY-1997; 97US-0047590.  
 PR 23-MAY-1997; 97US-0047592.  
 PR 23-MAY-1997; 97US-0047593.  
 PR 23-MAY-1997; 97US-0047594.  
 PR 23-MAY-1997; 97US-0047595.  
 PR 23-MAY-1997; 97US-0047596.  
 PR 23-MAY-1997; 97US-0047597.  
 PR 23-MAY-1997; 97US-0047598.  
 PR 23-MAY-1997; 97US-0047599.  
 PR 23-MAY-1997; 97US-0047600.  
 PR 23-MAY-1997; 97US-0047601.  
 PR 23-MAY-1997; 97US-0047612.  
 PR 23-MAY-1997; 97US-0047613.  
 PR 23-MAY-1997; 97US-0047614.  
 PR 23-MAY-1997; 97US-0047615.  
 PR 23-MAY-1997; 97US-0047616.  
 PR 23-MAY-1997; 97US-0047617.  
 PR 23-MAY-1997; 97US-0047618.  
 PR 23-MAY-1997; 97US-0047632.  
 PR 23-MAY-1997; 97US-0047633.  
 PR 06-JUN-1997; 97US-0048964.  
 PR 06-JUN-1997; 97US-0048974.  
 PR 13-JUN-1997; 97US-0049610.  
 PR 08-JUL-1997; 97US-0051926.  
 PR 16-JUL-1997; 97US-0052874.  
 PR 18-AUG-1997; 97US-0055724.  
 PR 22-AUG-1997; 97US-0056630.  
 PR 22-AUG-1997; 97US-0056631.  
 PR 22-AUG-1997; 97US-0056632.  
 PR 22-AUG-1997; 97US-0056636.  
 PR 22-AUG-1997; 97US-0056637.  
 PR 22-AUG-1997; 97US-0056662.  
 PR 22-AUG-1997; 97US-0056664.  
 PR 22-AUG-1997; 97US-0056845.  
 PR 22-AUG-1997; 97US-0056862.  
 PR 22-AUG-1997; 97US-0056864.  
 PR 22-AUG-1997; 97US-0056872.  
 PR 22-AUG-1997; 97US-0056874.  
 PR 22-AUG-1997; 97US-0056875.  
 PR 22-AUG-1997; 97US-0056876.  
 PR 22-AUG-1997; 97US-0056877.  
 PR 22-AUG-1997; 97US-0056878.  
 PR 22-AUG-1997; 97US-0056879.  
 PR 22-AUG-1997; 97US-0056880.  
 PR 22-AUG-1997; 97US-0056881.  
 PR 22-AUG-1997; 97US-0056882.  
 PR 22-AUG-1997; 97US-0056884.  
 PR 22-AUG-1997; 97US-0056886.  
 PR 22-AUG-1997; 97US-0056887.  
 PR 22-AUG-1997; 97US-0056888.  
 PR 22-AUG-1997; 97US-0056889.  
 PR 22-AUG-1997; 97US-0056892.  
 PR 22-AUG-1997; 97US-0056893.  
 PR 22-AUG-1997; 97US-0056894.  
 PR 22-AUG-1997; 97US-0056903.  
 PR 22-AUG-1997; 97US-0056908.  
 PR 22-AUG-1997; 97US-0056909.  
 PR 22-AUG-1997; 97US-0056910.  
 PR 22-AUG-1997; 97US-0056911.  
 PR 05-SEP-1997; 97US-0057650.  
 PR 05-SEP-1997; 97US-0057659.  
 PR 05-SEP-1997; 97US-0057761.  
 PR 12-SEP-1997; 97US-0058785.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;  
 PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;  
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 XX  
 DR WPI: 1998-506364/43.  
 P-PSDB: AAW74777.

XX  
 PR New isolated human genes and the secreted polypeptide(s) they encode  
 PR useful for diagnosis and treatment of e.g. cancers, neurological  
 PR disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS Claim 1: Page 279-280; 721pp; English.  
 XX  
 CC This sequence represents a nucleic acid molecule designated Gene 47 from  
 CC the human CDNA clone HOGAV75 (deposited as clone ATCC 97899 and ATCC  
 CC 209045) which encodes a secreted human protein. The gene can be used to  
 CC generate fusion proteins by linking to the gene to a human  
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of  
 CC the fused protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W5026)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new  
 CC polynucleotides. Specific uses are described for each of the 186  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAV59511 for described uses).  
 XX

SQ Sequence 1220 BP; 382 A; 207 C; 258 G; 372 T; 1 other;

alignment\_scores:                      length:    101  
     quality:    5.168                      gaps:       0  
 Percent Similarity: 100.000            Percent Identity: 99.010

alignment\_block:

US-09-484-964-2 x AAV59557 ..

Align seq 1/1 to: AAV59557 from: 1 to: 1220

1 MetSerAspGlnGluAlaLysProSerThrGluAspGluAspLysLys 17  
 128 ATGCTGACCCAGGAGCAAAACCTTCACCTGAGGACTTGGGGATAGAA 177  
 17 GlnGlyGlyTyrTleLysLeuLysValIleGlyGlnAspSerSerLys 34  
 178 GGAAGTGAATATATTAACCTCAAGTCATTGACAGATAGCAGTGA 227  
 34 IeHisPheLysValLysMetThrThrHisLeuLysLysLysGluSer 50  
 228 TTCACCTCAAGTCAAAATGACAAACATCTCAAGAAACCAAGATCA 277  
 51 TyrCysGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuPhe 67  
 278 TACGTCAAGACAGCGGTGTCCATCAATCAATCACTCAGGTTCTTGA 327  
 67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84  
 328 GGGTCAGAGAAATGCTGATATCACTACTCCAAAGAACTGGAAATGAGG 377  
 84 LysGluAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThr 100  
 378 AAGAGATGTGATTGAAGTTATCAGGACAAACGGGGGTCATTCAACA 427  
 101 Val 101  
 428 GTT 430

seq\_name: /STD1/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAZ35859

seq\_documentation\_block:

ID AAZ35859 standard; cDNA: 1514 BP.

AC AAZ35859;

DT 03-FEB-2000 (first entry)

DE Human sentrin nucleotide sequence.  
 XX  
 KW Human; sentrin; antisense oligonucleotide; phosphorothioate;  
 XX inhibition; modulation; expression; diagnosis; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 136..441  
 FT /tag= a  
 FT /product= "sentrin"  
 XX  
 PN US5985664-A.  
 XX  
 PD 16-NOV-1999.  
 XX  
 PF 17-DEC-1998; 98US-0213768.  
 XX  
 PR 17-DEC-1998; 98US-0213768.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 PI Baker BF, Cowser LM;  
 XX  
 DR WPI: 2000-022284/02.  
 DR P-PSDB: AA149967.  
 XX  
 PT Antisense compound which modulates human sentrin expression, useful for  
 XX treating diseases associated with sentrin expression -  
 XX  
 PS Example 13; Column 39-42; 29pp; English.  
 XX  
 CC The present invention describes an antisense compound (I) 8-30  
 CC nucleotides long targeted to a nucleic acid molecule encoding human  
 CC sentrin. The antisense compound comprises a phosphorothioate antisense  
 CC oligonucleotide which inhibits expression of human sentrin. (I) is  
 CC useful for inhibiting expression of sentrin in human cells or tissues  
 CC in vitro, for treating humans or other animals suspected of having or  
 CC being prone to a disease associated with sentrin expression. (I) can  
 CC also be used for research or diagnostic purposes. The present  
 CC sequence encodes human sentrin.  
 XX  
 SQ Sequence 1514 BP; 462 A; 256 C; 314 G; 482 T; 0 other;

alignment\_scores:  
 Quality: 522.00 Length: 101  
 Ratio: 5.168 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
 US-09-484-964-2 x AA235859 ..

Align seg 1/1 to: AA235859 from: 1 to: 1514

```

1 MeterspGingluAlaLysProSerThGluaspLeuGlyAspLys 17
|||||
136 ATGCTGCGAGGAGGCAAACTTCAACTGAGGAGGATTAAGAA 185
17 sGluGluValTyrTleLysLeuValIleGluGluAspSerGlu 34
|:::|
186 GGAAGGGAATATATTAACCAAGTCATTGGACAGATAGCGTGA 235
34 IeHisPheValLysMetThrThrHisLeuLysLysLeuGluSer 50
|||||
236 TTCACCTTCAAGTGAATGACACACATCTCAGAAACTCAAGATCA 285
51 TyrCysGluArgGluGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
286 TACTGTCAAGACAGGCGTTCACATGAAATCACTCAGGTTCTCTTGA 335
67 uGluGluArgTleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
|||||

```

```

336 GGGTCAGAGATTGCTGATATCATCTCCAAAGAACTGGGATGAGG 385
84 luGluaspValIleGluValTyrGluGluGlnThrGlyGlyHisSerThr 100
|||||
386 AAGAAAGATGTGATTGAAGTTTATCAGGAACAAACGGGGCTCATTCACA 435
101 Val 101
|||
436 GTT 438

```

seq\_name: /SIDSL/gcdata/geneseq/geneseqn/NA2000.DAT:AAA16215

seq\_documentation\_block:  
 ID AAA16215 standard; DNA; 616 BP.  
 XX  
 AC AAA16215;  
 XX  
 DT 14-JUN-2000 (first entry)  
 XX  
 DE Human colon cancer differentially expressed nucleotide sequence #220.  
 XX  
 KW Colon cancer; detect; differential expression; human; treatment;  
 KW detect mutation; non-invasive diagnostic method; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200012702-A2.  
 XX  
 PD 09-MAR-2000.  
 XX  
 PF 30-AUG-1999; 99WO-US19424.  
 XX  
 PR 31-AUG-1998; 98US-0098639.  
 PR 27-JAN-1999; 99US-0117393.  
 XX  
 PA (FARB ) BAYER CORP.  
 XX  
 PI Endege WO, Steimann KE, Astle JH, Burgess CC, Carroll E;  
 PI Carino TJ, Divedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;  
 PI Schlegel R;  
 XX  
 DR WPI: 2000-256641/22.  
 XX  
 PT Novel nucleic acids and proteins for identifying therapeutic agents  
 PT useful for treating and diagnosing cancer, especially colon cancer -  
 PS  
 PS Claim 16; Page 211-212; 345pp; English.  
 XX  
 CC This sequence represents a human nucleotide sequence which is  
 CC differentially expressed in colon cancer cells compared to the expression  
 CC levels in normal cells. The nucleotide sequence can be used as a source  
 CC of primers and probes. The nucleotide sequence is useful for determining  
 CC the phenotype of a cell by detecting the differential expression of the  
 CC sequence relative to a normal cell. The probes derived from the sequence  
 CC can also be used to determine the phenotype of cells in a sample. Probes  
 CC and antibodies which hybridise to the nucleotide sequence can also be  
 CC used to determine the phenotype of a cell. The primers are useful for  
 CC detecting a mutation in a test nucleotide sequence and also for detecting  
 CC cancer, preferably colon cancer. Antibodies against the protein encoded  
 CC by the nucleotide sequence can also be used in a method to detect colon  
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing  
 CC colon cancer at an early stage.  
 XX  
 SQ Sequence 616 BP; 181 A; 109 C; 149 G; 170 T; 7 other;

alignment\_scores:  
 Quality: 503.00 Length: 102  
 Ratio: 5.030 Gaps: 1  
 Percent Similarity: 98.039 Percent Identity: 97.059

alignment\_block:  
 US-09-484-964-2 x AAA16215 ..

Align seg 1/1 to: AAA16215 from: 1 to: 616

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
127 ATGCTGACCAAGGAGGCAAAACCTTCAACTGAGAGCTTGGGGGATTAAGAA 176
  |||
17 sGlnGlyGluTyrIleLeuLysValIleGlyGlnAspSerSerGlu 34
  |||
177 GGAAGGCAATATTTAACTCAAAAGTCATTGGACAGATAGCAGTGA 226
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLysGluSer 50
  |||
227 TTCACCTCAAGTGAATAATGACACACATCTCAAGAAACTCAAAAGATCA 276
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
  |||
277 TACGTCAAGACAGGGGTGTCCAAATGATGACTCAAGTTCTCTTGA 326
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
  |||
327 GGGTCACAGAAATTCGTATATCATCTCAAAAGAACTGGGAATGAGAG 376
  |||
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerTh 100
  |||
377 AAGAGATGTGATGAAAGTTTATCANGAACAAACCGGGGTCAATCAAC 426
  |||
100 rVal 101
  |||
427 AGTT 430

```

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:AA163339

seq\_documentation\_block:

ID AAT63339 standard; CDNA: 372 BP.

XX AAT63339;

DT 17-AUG-1997 (first entry)

DE Human host cell protein NPI-6 partial cDNA clone.

KW Host cell protein; NPI-6; nucleoprotein interactor 6;

KW Influenza virus; replication; antiviral; virucide; ss.

OS Homo sapiens.

PN W09712967-A1.

PD 10-APR-1997.

PF 06-OCT-1995; 95WO-US13044.

PR 06-OCT-1995; 95WO-US13044.

PA (MOON ) MOUNT SINAI MEDICAL CENT.

PI Oneill R, Palese P;

DR WPI; 1997-226211/20.

XX New isolated DNA which encodes viral interacting proteins - used in  
 PT assays to isolate products for inhibiting viral protein binding  
 PT which is required for infection, replication, assembly or release

PS Disclosure; Fig 11; 98pp; English.

XX CDNA clones (AAT63335-39) comprise partial sequences for human

CC nucleoprotein interactor proteins NPI-2 to NPI-6, respectively.

CC These are host cell proteins which interact with influenza virus

CC nucleoprotein (NP) and which may be accessory proteins required for

CC influenza virus replication. NPI-6 was identified as a novel

CC protein sequence. NPI sequences were isolated by interactive trap

CC selection using LexA-NP as bait and yeast transformed with an HeLa  
 CC CDNA library. NPI-1 (see also AAT63334) and NPI-2 to NPI-6 cDNAs were  
 CC cloned and analysed. The clones can be used to produce human NPI  
 CC proteins for use in identifying cDNAs that inhibit viral  
 CC replication.

SQ Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

alignment\_scores:

Quality: 492.00 Length: 97

Ratio: 5.125 Gaps: 0

Percent Similarity: 98.969 Percent Identity: 97.938

alignment\_block:

US-09-484-964-2 x AAT63339 ..

Align seg 1/1 to: AAT63339 from: 1 to: 372

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
80 ATGCTGACCAAGGAGGCAAAACCTTCAACTGAGAGCTTGGGGGATTAAGAA 129
  |||
17 sGlnGlyGluTyrIleLeuLysValIleGlyGlnAspSerSerGlu 34
  |||
130 GGAAGGCAATATTTAACTCAAAAGTCATTGGACAGATAGCAGTGA 179
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLysGluSer 50
  |||
180 TTCACCTCAAGTGAATAATGACACACATCTCAAGAAACTCAAAAGATCA 229
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
  |||
230 TACGTCAAGACAGGGGTGTCCAAATGATGACTCAAGTTCTCTTGA 279
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
  |||
280 GGGTCACAGAAATTCGTATATCATCTCAAAAGAACTGGGAATGAGAG 329
  |||
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerTh 97
  |||
330 AAGAGATGTGATGAAAGTTTATCANGAACAAACCGGGGT 370

```

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF59399

seq\_documentation\_block:

ID AAF59399 standard; DNA: 372 BP.

XX AAF59399;

DT 02-MAY-2001 (first entry)

DE Human host cell protein NPI-6 partial nucleotide sequence SEQ ID NO:11.

KW Identification; antiviral; viral protein; viral replication; NP;

KW Viral infection; nucleoprotein; ds.

OS Homo sapiens.

PN W0200111335-A2.

PD 15-FEB-2001.

PF 11-AUG-2000; 2000MO-US22257.

PR 11-AUG-1999; 99US-0148263.

PA (MOON ) MOUNT SINAI SCHOOL MEDICINE.

PI O'Neill R, Harty R, Palese PM;

DR WPI; 2001-168816/17.

PT Identifying a substance that inhibits the interaction between a viral  
 PT protein and a host cell protein, useful for the discovery of new  
 PT antiviral compounds -

PS Disclosure; Fig 11; 147pp; English.

CC The present invention describes a method (M1) for identifying a  
 CC substance that inhibits the interaction of a viral protein (VP) with a  
 CC host cell protein (HP). The method comprises: (a) contacting HP with VP  
 CC in the presence of a test substance; and (b) detecting complex formation,  
 CC where the ability of the test substance to inhibit HP/VP interaction is  
 CC indicated by a decrease in complex formation. The antiviral compounds  
 CC that inhibit the interaction between a host protein (NS1-BP or NP1-1)  
 CC and a viral protein (NS1) are useful for treating or inhibiting viral  
 CC infection, preferably influenza and rhinovirus infection, in humans.  
 CC Antiviral compounds include peptides and antibodies. In particular  
 CC compositions comprising a polypeptide containing an amino acid sequence  
 CC corresponding to the NP-NS1 domain of the influenza virus NP protein,  
 CC which inhibits the specific interaction of the influenza virus NP protein  
 CC with the influenza virus NP protein are useful for treating or inhibiting  
 CC human host cell infection in humans. The present sequence represents a  
 CC human host cell protein NP1-6 partial nucleotide sequence, which is used  
 CC in the exemplification of the present invention.

SQ Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

#### Alignment\_scores:

Quality: 492.00 Length: 97  
 Ratio: 5.125 Gaps: 0  
 Percent Similarity: 98.969 Percent Identity: 97.938

#### alignment\_block:

US-09-484-964-2 x AAF59399 ..

Align seg 1/1 to: AAF59399 from: 1 to: 372

```

1 MetSerAspGlnGlnAlaLysProSerThrgluAspLeuGlyAspLysLy 17
|||||
80 ATGCTGACACAGAGGAGCAAAACCTTCACACTGAGGAGTGGGAGTAAGAA 129
|||||
17 sGlnGlyGluTrpIleLysLeuLysValIleGlyGlnAspSerSerglu 34
|:::|
130 GGAAGGTGAATATATTAACCTCAAGTCATGACAGAGTAGCAGTGAGA 179
|||||
34 IeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
|||||
180 TTCACCTCAAGTGAATGACAAACACATCTCAAGAAACCTCAAGAAATCA 229
|||||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
230 TACTGTCAAAAGACAGGGGTCTCCAAATGATCACTCAAGTTCTCTTTGA 279
|||||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
280 GGGTCAGAGAAATTCCTGATTAATCACTCCAAAAGAACTGGGAATGAGG 329
|||||
84 IuGluAspValIleGluValTyrGlnGluGlnThrGlyGly 97
|||||
330 AAGAAGTTGATGATGAACTTATTCAGAACAAACGGGGGCT 370

```

seq\_name: /STDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAAT03735

seq\_documentation\_block:

ID AAT03735 standard; cDNA; 425 BP.

AC AAT03735;

DT 26-MAR-1996 (first entry)

DE TNF-R p55IC/Fas-IC-binding protein DD11 cDNA clone.

XX Tumour necrosis factor receptor; TNF-R; p55IC; Fas-IC;

KW intracellular domain binding protein; ss.

XX Homo sapiens.

OS W09531544-A1.

PN 23-NOV-1995.

PD 11-MAY-1995.

PF 95WO-US05854.

PR 02-OCT-1994; 94II-0111125.

PR 11-MAY-1994; 94II-0109632.

XX (WEIN/) WEINMURZEL H.

PA (YEDA ) YEDA RES & DEV CO LTD.

PI Boldin M, Mett I, Varfolomeev E, Wallach D;

DR WPI; 1996-010930/01.

XX TNF-NGF receptor superfamily intracellular domain-binding proteins -

XX or HIV-infected cells

XX Claim 18; Fig 12; 96pp; English.

XX cDNA clone DD11 (AAT03735) was isolated by screening a human HeLa

XX library using the human tumour necrosis factor receptor (TNF-R)

XX p55 death domain (p55DD) as 'bait'. The full length of the DD11

XX transcript is approx 1.2 kb. The protein encoded by cDNA DD11

XX interacts strongly with p55DD (amino acids 326-414) and also with

XX human and mouse FAS intracellular domain. Such proteins may

XX be used modulate TNF-R function.

SQ Sequence 425 BP; 135 A; 80 C; 115 G; 95 T; 0 other;

#### alignment\_scores:

Quality: 470.50 Length: 102  
 Ratio: 4.801 Gaps: 2  
 Percent Similarity: 96.078 Percent Identity: 91.176

#### alignment\_block:

US-09-484-964-2 x AAT03735 ..

Align seg 1/1 to: AAT03735 from: 1 to: 425

```

1 MetSerAspGlnGlnAlaLysProSerThrgluAspLeuGlyAspLysL 17
|||||
118 ATGCTGACACAGAGGAGCAAAACCTTCACACTGAGGAGTGGGAGTAAGAA 167
|||||
17 ysGlnGlyGluTrpIleLysLeuLysValIleGlyGlnAspSerSerglu 33
|:::|
168 AGAAGGTGAATATTAACCTCAAGTCATGACAGAGTAGCAGTAGAG 217
|||||
34 IeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSe 50
|||||
218 ATTCACCTCAAGTGAATGACAAACACATCTCAAGAAACCTCAAGAAATC 267
|||||
50 rTyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
268 ATACTGTCAAAAGACAGGGGTCTCCAAATGATCACTCAAGTTCTCTTTG 317
|||||
67 IuGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 83
|||||
318 AGGGTCAGAGAAATTCCTGATTAATCACTCCAAAAGAACTGGGAATGAG 367
|||||
84 GluGluAspVal...IleGluValTyrGlnGluGlnThrGlyGlyHisSe 99
|:::|
368 AAGAAGATTGATGATGAACTTATTCAGAACAAACGGGGGCTCATTC 417
|||||
99 rThr 100
|||||

```



PR 21-APR-1999; 99US-0130449  
PR 23-APR-1999; 99US-0130510  
PR 23-APR-1999; 99US-0130891  
PR 28-APR-1999; 99US-0131449  
PR 30-APR-1999; 99US-0132048  
PR 30-APR-1999; 99US-0132407  
PR 04-MAY-1999; 99US-0132484  
PR 05-MAY-1999; 99US-0132486  
PR 06-MAY-1999; 99US-0132486  
PR 06-MAY-1999; 99US-0132487  
PR 07-MAY-1999; 99US-0132863  
PR 11-MAY-1999; 99US-0134256  
PR 14-MAY-1999; 99US-0134218  
PR 14-MAY-1999; 99US-0134219  
PR 14-MAY-1999; 99US-0134221  
PR 14-MAY-1999; 99US-0134370  
PR 18-MAY-1999; 99US-0134768  
PR 19-MAY-1999; 99US-0134941  
PR 20-MAY-1999; 99US-0135124  
PR 21-MAY-1999; 99US-0135353  
PR 24-MAY-1999; 99US-0135629  
PR 25-MAY-1999; 99US-0136021  
PR 27-MAY-1999; 99US-0136392  
PR 28-MAY-1999; 99US-0136782  
PR 01-JUN-1999; 99US-0137222  
PR 03-JUN-1999; 99US-0137528  
PR 04-JUN-1999; 99US-0137502  
PR 07-JUN-1999; 99US-0137724  
PR 08-JUN-1999; 99US-0138094  
PR 10-JUN-1999; 99US-0138540  
PR 10-JUN-1999; 99US-0138847  
PR 14-JUN-1999; 99US-0139119  
PR 16-JUN-1999; 99US-0139452  
PR 16-JUN-1999; 99US-0139453  
PR 17-JUN-1999; 99US-0139492  
PR 18-JUN-1999; 99US-0139454  
PR 18-JUN-1999; 99US-0139455  
PR 18-JUN-1999; 99US-0139456  
PR 18-JUN-1999; 99US-0139457  
PR 18-JUN-1999; 99US-0139458  
PR 18-JUN-1999; 99US-0139459  
PR 18-JUN-1999; 99US-0139460  
PR 18-JUN-1999; 99US-0139461  
PR 18-JUN-1999; 99US-0139462  
PR 18-JUN-1999; 99US-0139463  
PR 18-JUN-1999; 99US-0139750  
PR 18-JUN-1999; 99US-0139763  
PR 21-JUN-1999; 99US-0139817  
PR 22-JUN-1999; 99US-0139899  
PR 23-JUN-1999; 99US-0140353  
PR 23-JUN-1999; 99US-0140354  
PR 24-JUN-1999; 99US-0140695  
PR 28-JUN-1999; 99US-0140823  
PR 29-JUN-1999; 99US-0140991  
PR 30-JUN-1999; 99US-0141287  
PR 01-JUL-1999; 99US-0141842  
PR 01-JUL-1999; 99US-0142154  
PR 02-JUL-1999; 99US-0142055  
PR 06-JUL-1999; 99US-0142390  
PR 08-JUL-1999; 99US-0142803  
PR 09-JUL-1999; 99US-0142920  
PR 12-JUL-1999; 99US-0142877  
PR 13-JUL-1999; 99US-0143342  
PR 14-JUL-1999; 99US-0143624  
PR 15-JUL-1999; 99US-0144005  
PR 16-JUL-1999; 99US-0144085  
PR 16-JUL-1999; 99US-0144086  
PR 19-JUL-1999; 99US-0144325  
PR 19-JUL-1999; 99US-0144331  
PR 19-JUL-1999; 99US-0144332  
PR 19-JUL-1999; 99US-0144333  
PR 19-JUL-1999; 99US-0144334  
PR 19-JUL-1999; 99US-0144335

---

PR 20-JUL-1999; 99US-0144352  
PR 20-JUL-1999; 99US-0144632  
PR 20-JUL-1999; 99US-0144884  
PR 21-JUL-1999; 99US-0144814  
PR 21-JUL-1999; 99US-0145086  
PR 21-JUL-1999; 99US-0145088  
PR 22-JUL-1999; 99US-0145087  
PR 22-JUL-1999; 99US-0145087  
PR 22-JUL-1999; 99US-0145089  
PR 22-JUL-1999; 99US-0145192  
PR 23-JUL-1999; 99US-0145145  
PR 23-JUL-1999; 99US-0145224  
PR 23-JUL-1999; 99US-0145224  
PR 26-JUL-1999; 99US-0145276  
PR 27-JUL-1999; 99US-0145913  
PR 27-JUL-1999; 99US-0145918  
PR 27-JUL-1999; 99US-0145919  
PR 28-JUL-1999; 99US-0145951  
PR 02-AUG-1999; 99US-0146386  
PR 02-AUG-1999; 99US-0146388  
PR 02-AUG-1999; 99US-0146389  
PR 03-AUG-1999; 99US-0147038  
PR 04-AUG-1999; 99US-0147204  
PR 04-AUG-1999; 99US-0147302  
PR 05-AUG-1999; 99US-0147192  
PR 05-AUG-1999; 99US-0147260  
PR 06-AUG-1999; 99US-0147303  
PR 06-AUG-1999; 99US-0147416  
PR 09-AUG-1999; 99US-0147493  
PR 09-AUG-1999; 99US-0147935  
PR 10-AUG-1999; 99US-0148171  
PR 11-AUG-1999; 99US-0148319  
PR 12-AUG-1999; 99US-0148341  
PR 13-AUG-1999; 99US-0148565  
PR 13-AUG-1999; 99US-0148684  
PR 16-AUG-1999; 99US-0149368  
PR 17-AUG-1999; 99US-0149175  
PR 18-AUG-1999; 99US-0149426  
PR 20-AUG-1999; 99US-0149722  
PR 20-AUG-1999; 99US-0149723  
PR 20-AUG-1999; 99US-0149929  
PR 23-AUG-1999; 99US-0149902  
PR 23-AUG-1999; 99US-0149930  
PR 25-AUG-1999; 99US-0150566  
PR 26-AUG-1999; 99US-0150884  
PR 27-AUG-1999; 99US-0151065  
PR 27-AUG-1999; 99US-0151066  
PR 27-AUG-1999; 99US-0151080  
PR 27-AUG-1999; 99US-0151303  
PR 30-AUG-1999; 99US-0151303  
PR 31-AUG-1999; 99US-0151438  
PR 01-SEP-1999; 99US-0151930  
PR 07-SEP-1999; 99US-0152363  
PR 10-SEP-1999; 99US-0153070  
PR 13-SEP-1999; 99US-0153758  
PR 15-SEP-1999; 99US-0154018  
PR 16-SEP-1999; 99US-0154039  
PR 20-SEP-1999; 99US-0154779  
PR 22-SEP-1999; 99US-0155139  
PR 23-SEP-1999; 99US-0155486  
PR 24-SEP-1999; 99US-0155659  
PR 28-SEP-1999; 99US-0156458  
PR 29-SEP-1999; 99US-0156596  
PR 04-OCT-1999; 99US-0157117  
PR 05-OCT-1999; 99US-0157753  
PR 06-OCT-1999; 99US-0157865  
PR 07-OCT-1999; 99US-0158029  
PR 08-OCT-1999; 99US-0158232  
PR 12-OCT-1999; 99US-0158369  
PR 13-OCT-1999; 99US-0159293  
PR 13-OCT-1999; 99US-0159294  
PR 13-OCT-1999; 99US-0159295  
PR 14-OCT-1999; 99US-0159329  
PR 14-OCT-1999; 99US-0159330





seq\_documentation\_block:  
ID AAC35316 Standard; DNA; 643 bp.  
AC AAC35316;  
XX  
XX  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9746.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143942.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144881.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.



PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140623.  
PR 29-JUN-1999; 99US-0140695.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142927.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145513.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150984.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0155658.  
PR 29-SEP-1999; 99US-0155659.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157823.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160771.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Quality:	221.00	Length:	85
Ratio:	3.250	Gaps:	1
Percent Similarity:	80.000	Percent Identity:	51.765

alignment\_block;  
ITS-09-484-96A-2 x AAC48770

Align seg 1/1 to: AAC48770 from: 1 to: 6339

15 AsplryLys .....GInGIoLnuYrILLeYsLeuValILeLYcI 29  
 172 GACACAGAGCCTGCACCAAGGAGCCTACATCATCTCAAGTCAAGGGACA 221  
 29 nAsPSeRSeRgIuLeILeILeISpHeLySValLYsMeThrThrILeILeYsL 46  
 222 GGAAGGATATGAAGCTCTTGTAGGATCAAGAGAAAGACCTACACTAAAA 271  
 46 ySLLeuLYsGLuSeRfTyTcYsGInATcRGInLYValPrOmEtAsnSerLeu 62  
 272 AACGcATGcATcATcCTTAcTcTcTcGAcCCTcAcTcTcTcGATTTCAAcTcAT 321  
 63 ArGPHeuLeuPHeGLuGLyGInATgILeALeAsPAsnHisThrProLYsGL 79  
 322 GCTTCTCTGTTGAAGGCGTGTGCTTCTCGCGGACAGACTCCAGATGA 371  
 79 uLeuGLyMeTcGLuGInLYAsPValILeGLuValTYrGInGLuGInThrg 96  
 372 GCTTGAATGTGAAGATGAGATGAGATCGATGCATGCATTCATCAGACTG 421  
 96 LyGLy 97  
 122 GNGGT 426

seq\_name: /SIDS1/gcdata/geneseq/geneseqn/NA2000.DAT:AAA31186

```
seq_documentation_block:
ID   AAA31186 standard; DNA; 351 BP.
```

AC	AAA31186;	
XX		
DT	05-JUL-2000	(first entry)
XX		
DE	Plant microsatellite marker #147.	
XX		
KW	Plant microsatellite sequence; core repeat sequence; detection; probe	
KW	DNA polymorphism; genome mapping; physical mapping; fingerprinting;	
XX	variety identification; genetic variability evaluation; primer; ss.	
XX		
OS	Eucalyptus grandis.	
XX		
PN	WO967421-A1.	
XX		
PD	29-DEC-1999.	
XX		
FE	25-JUN-1999;	99WO-NZ00092.
XX		
PR	25-JUN-1998;	9805-0105307.
XX		
PA	(GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.	
PA	(FLET-) FLETCHER CHALLENGE FORESTS LTD.	
XX		
PI	Havukkalehti, Bloksberg LN, Glenn M;	
XX		
DR	WPI; 2000-116958/10.	
PT	New plant microsatellite markers and associated flanking species for	
PT	the detection of polymorphic genetic markers -	
XX		
PS	Claim 1; Page 119; 392pp; English.	

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences  
CC and associated flanking species. The sequences comprise a central core  
CC repeat sequence, especially selected from the sequences AAA32094-A32096  
CC with left and right flanking sequences. The polymorphic sequences  
CC can be used in the detection of DNA polymorphisms, in genome mapping,  
CC in physical mapping, in positional cloning of genes, in variety  
CC identification and in evaluation of genetic variability within and  
CC between plant tissues, populations, cultivars, species and species  
CC groups. They may also be used to design hybridization probes for  
CC oligonucleotide fingerprinting and library screening and to design  
CC primers for microsatellite-primed PCR. Microsatellite markers are  
CC useful to locate specific economically useful genes in plant genomes.  
XX  
S0 Sequence 351 BP; 97 A; 69 C; 112 G; 73 T; 0 other;

XX Sequence 351 BP; 97 A; 69 C; 112 G; 73 T; 0 other;  
SQ

[illegible]

```
alignment_block:
US-09-484-964-2 x AAA31186
```

Align seg 1/1 to: AAA31186 from: 1 to: 351

```

15 AspLyLyysGInGlyGlu..... TyrIleLyysLeuysValIleGI 28
   |||||..... |||||..... |||||..... |||||..... ||
78 GACAAACAAACCCCGGATGAGCCGCCCAACATCATCTGAAAGCAAGG 127
   |||||..... |||||..... |||||..... |||||..... ||
28 yGlnaSPserSergIuIleHsPheLyysValLyysMethrThriSteul 45
   |||||..... |||||..... |||||..... |||||..... ||
128 CCGAGATGCTAATGAGCTCTTTTTCAGATCAACGAAGCACCCAGTTGA 177
   |||||..... |||||..... |||||..... |||||..... ||
45 yslYsLeuLyysGlnuSerTYrCYsGlnATrGInGlyAlaProMeCAsSer 61
   |||||..... |||||..... |||||..... |||||..... ||
178 GGAACACTCATGACCGGCTACTGCATAGGCAATCTGTGAGATTAACTCC 227
   |||||..... |||||..... |||||..... |||||..... ||
62 LeuATrPheLeuPheGluLyGlnAlaGlyIleAlaSPasnHsIstrProLy 78
   :::: |||||..... |||||..... |||||..... |||||..... ||
228 ATTGCTTTCTTAATTCATGGGGCTGACTCTCGTGTGACAGACTCTGA 277
   |||||..... |||||..... |||||..... |||||..... ||
78 sGluLeuGlyMetGluGlnGluAlaSPValIleGluValTyGInGluGI 95
   |||||..... |||||..... |||||..... |||||..... ||
278 TGAAGCTCGAGATGAGAGATGTGATGAATAATGCAGCCATCTCCACAAA 327
   |||||..... |||||..... |||||..... |||||..... ||
95 hrcIyGlyIhIs 98
   |||||..... |||||..... |||||..... |||||..... ||
128 CTGGGGGATAC 338

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC46716

```
seq_documentation_block:
ID   AAC46716 standard; DNA; 505 BP.
```

AA	AAC46716;
AC	
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Zea mays DNA fragment SEQ ID NO: 51143.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control;
KW	Protein identification; signal transduction pathway; metabolic
KW	pathway; promoter; termination sequence; corn; ss.
XX	
OS	Zea mays subsp. mays.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0122548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132048.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134766.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138099.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140635.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149175.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.





```

APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/853,974
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0289 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 473 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGFET03
CLONE: 2361410
US-08-853-974-2

alignment_scores:
Quality: 217.00 Length: 98
Ratio: 2.893 Gaps: 2
Percent Similarity: 76.531 Percent Identity: 44.898

alignment_block:
US-09-484-964-2 x US-08-853-974-2
..
Align seg 1/1 to: US-08-853-974-2 from: 1 to: 473

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLy 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 ATGTCCGAGGAGAACCCCAAGAG.....GGTGTCAAGAC 118

17 sGlnGlyIuTyriLeuLysLeuLysValIleGlyGlnAspSerSerGluI 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 AGAGATATACCAATCACTGAAGGTGGCCGGGAGAGAGCGCTCCGTGG 168

34 leHisPheValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 TGCAGTTCAAGATCAAGCAGCAGCGCGGTGAGCAGCAAGCTGATGAAGGCC 218

51 TyrcysGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheG1 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 TACGCCAGAGAGCGAGGCTTGTCAATGAGCAGATCAGATTCAGATTGCGA 268

67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 CGGGCAGCCCAATGATGAAGTGAAGTCCAGACAGCAGCTGAGAGTAGAGG 318

84 lUGluAspValIleGluValTyrcysGlnGlnGlnGlnGlnGlnGlnGly 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

319 ACGAGACACCATGACGCTGTTTCCAAAGCAGACAGCGGAGGT 362
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-172-988-2
seq_documentation_block:
; Sequence 2, Application US/09172988
; Patent No. 6136935
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/172,988
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/853,974
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0289 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGFET03
; CLONE: 2361410
US-09-172-988-2

alignment_scores:
Quality: 217.00 Length: 98
Ratio: 2.893 Gaps: 2
Percent Similarity: 76.531 Percent Identity: 44.898

alignment_block:
US-09-484-964-2 x US-09-172-988-2
..
Align seg 1/1 to: US-09-172-988-2 from: 1 to: 473

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLy 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 ATGTCCGAGGAGAACCCCAAGAG.....GGTGTCAAGAC 118

17 sGlnGlyIuTyriLeuLysLeuLysValIleGlyGlnAspSerSerGluI 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 AGAGATATACCAATCACTGAAGGTGGCCGGGAGAGAGCGCTCCGTGG 168

34 leHisPheValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 TGCAGTTCAAGATCAAGCAGCAGCGCGGTGAGCAGCAAGCTGATGAAGGCC 218

51 TyrcysGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheG1 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



Thu Aug 16 08:48:54 2001

us-09-484-964-2.rn1

Page 3

```
219 TACTGCAGAGCAGGCGCTGTTCATGAGCAGATCATTCAGTTGCA 268
67 uGlyGlnArgIleAlaSpasnhIsthPrPolySgIuLeuGlyMetGlu 84
   ||||| ||| :|||: ||||| :|||: |||||:
269 CGGCACCAATCATGAACTGACACTCCAGCAGCTGGAGATGGAGG 318
84 IuGluAspValIleGluValTyr...GlnGluGlnThrGlyGly 97
   :|||:|||||:|||||:|||||:|||||:|||||:
319 ACGGAGACCATCGACGTGTTCCTTCCAGCAGCAGAGAGGAGGT 362

seq_name: /cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-755-584-1

seq_documentation_block:
; Sequence 1, Application US/08755584
; Patent No. 5858711
; GENERAL INFORMATION:
; APPLICANT: Gilmcher, Laurie H.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..1248
; US-08-755-584-1

alignment_scores:
Quality: 101.00 Length: 102
Ratio: 1.656 Gaps: 2
Percent Similarity: 59.804 Percent Identity: 25.490

alignment_block:
US-09-484-964-2 x US-08-755-584-1 ..
Align seg 1/1 to: US-08-755-584-1 from: 1 to: 1946
6 AlalysProSerThrGluAspLeuGly..... 14
||| ||||| ||||| |||||
940 GCCACCCCTAGTACCCTTAAGCTTGGAGTGGCTGACATCATTCATGTTGTCT 989
```

```
15 .....AspLysLysGlnGlyIuTyrIleLysL 24
   :|||:|||||:|||||:|||||:|||||:
990 GGTCTGACAGAGCTCTTCAGAGGCCACAGACATCCCGAGAGCTCCGGC 1039
24 euLysValIleGlyGlnAspSerSerGlu...IleHisPheLysValLys 39
   ||||| ||| :|||: ||||| :|||: |||||:
1040 TCCGGGTGCGAGGGGAGAGAACACCAACCATGTTGGAGATCTCAGCTGT 1089
40 MetThrHisIleLysLysLysLysGluSerTyrCysGlnArgGlnG 56
   :|||:|||||:|||||:|||||:|||||:
1090 CCTGATCTCTCTTTAAGTTTCATGTCACACTATGAGAGCATGGG 1139
56 ValProMetAsnSerLeuArgPheLeuPheGluGlyGlnArgIleAla 73
   ||| :|||: ||||| :|||: |||||:
1140 ACTCTGTGACACAGACTCTCTCTTTGATGGAGCAACAGCTTTCAG 1189
73 spAsnhIsthPrPolySgIuLeuGlyMetGluGluGluAspValIleGlu 89
   :|||: ||| :|||:|||||:|||||:|||||:
1190 GCAAGAGAGCTGCAGCTGATCTGGGCTGGAAATCCGAGATCTCATGAA 1239
90 ValTyr 91
   |||||
1240 GTCTGG 1245

seq_name: /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-192-611-1

seq_documentation_block:
; Sequence 1, Application US/09192611
; Patent No. 6090561
; GENERAL INFORMATION:
; APPLICANT: Gilmcher, Laurie H.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,611
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..1248
; US-09-192-611-1
```



```

13 LeuGlyAspLysGlnGlyGluTyrIleLysLeuLysValIleGly1 29
|||||
2652 CTGGTCTCGCCCTCAGGAGGAGCAGATCGATCTTCGTGA...GACC 2698
29 naspserSerGluIleHis.PheLysValLysMetThrThrHisLeuLys 45
|||
2699 TGACCGGCAAGACTATCACCCTCGAGGTGAGTCTTCAGACACCATCGAC 2748
46 LysLeuLysGluSerTyrCysGlnArgGlnGlyValProMetAsnSerle 62
|||||
2749 AATGTCAAGGCCAAGATCCAGACAGAGGCGATCCACCGGACCA 2798
62 uArgPheLeuPheGlnGlyGlnArgIleAlaAspAsnHisThrProLys 79
|||||
2799 GCGTTGTCTTCCTGCGCAAGCAGCTGAGAGATGCCCGCACCTTCGCG 2848
79 LuLeuGlyMetGluGluGluAspValIleGluValTyrGlnGlnGluThr 95
::
2849 ATTACAACATCCAGAAAGAGACCTCCAGCTGTGCTCTCCGTTCAGG 2898
96 GlyGly 97
|||||
2899 GGTGGT 2904

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-462-092-1

seq_documentation_block:
; Sequence 1, Application US/08462092
; Patent No. 5614399
;
; GENERAL INFORMATION:
; APPLICANT: Peter H. Quail
; APPLICANT: Alan H. Christensen
; APPLICANT: Howard P. Hershey
; APPLICANT: Robert A. Sharrock
; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,092
; FILING DATE: 5-JUNE-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: 08/462,092
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-375-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3840 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1993..3591
;
; US-08-462-092-1

```

```

alignment_scores:
  quality: 78.50      length: 86
  ratio: 1.402      gaps: 2
  percent similarity: 65.116      percent identity: 22.093

alignment_block:
US-09-484-964-2 x US-08-462-092-1 ..
Align seg 1/1 to: US-08-462-092-1 from: 1 to: 3840

13 LeuGlyAspLysGlnGlyGluTyrIleLysLeuLysValIleGly1 29
|||||
2424 CTGGTCTCGCCCTCAGGAGGAGCAGATCTTCGTGA...GACC 2470
29 naspserSerGluIleHis.PheLysValLysMetThrThrHisLeuLys 45
|||
2471 TGACCGGCAAGACTATCACCCTCGAGGTGAGTCTTCAGACACCATCGAC 2520
46 LysLeuLysGluSerTyrCysGlnArgGlnGlyValProMetAsnSerle 62
|||||
2521 AAGCTCAAGGCCAAGATCCAGACAGAGGCGATCCCCACAGACCA 2570
62 uArgPheLeuPheGlnGlyGlnArgIleAlaAspAsnHisThrProLys 79
|||||
2571 GCGGTCTATCTTGTCTGGAAGCAGCTTGAGAGACGCGCGCTTCGCG 2620
79 LuLeuGlyMetGluGluGluAspValIleGluValTyrGlnGlnGluThr 95
::
2621 ACTACAACATCCAGAAAGAGACCTCCAGCTGTGCTCTCGCCCTCAGG 2670
96 GlyGly 97
|||||
2671 GGAGG 2676

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-746-822-1

seq_documentation_block:
; Sequence 1, Application US/08746822
; Patent No. 6020190
;
; GENERAL INFORMATION:
; APPLICANT: Peter H. Quail
; APPLICANT: Alan H. Christensen
; APPLICANT: Howard P. Hershey
; APPLICANT: Robert A. Sharrock
; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,822
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,092
; FILING DATE: 5-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: 08/462,092
; TELECOMMUNICATION INFORMATION:

```

us-09-484-964-2.rni

GENERAL INFORMATION:  
APPLICANT: MASCARENHAS, DESMONT

Page 7

```

seq_name: /cgn2_6/plodata/2/ina/5B_COMB.seq:us-08-854-811-42
seq_documentation_block:
  Sequence 42, Application US/08854811
  Patent No. 5914254
  GENERAL INFORMATION:
    APPLICANT: Mascarenhas, Desmond
    APPLICANT: Zhang, Yang
    APPLICANT: Olson, Pamela S.
    APPLICANT: Olsen, David R.
    APPLICANT: Cohen, Pedro A.
    TITLE OF INVENTION: EXPRESSION OF FUSION POLYPEPTIDES
    TITLE OF INVENTION: TRANSPORTED OUT OF THE CYTOPLASM WITHOUT LEADER
    TITLE OF INVENTION: SEQUENCES
    NUMBER OF SEQUENCES: 49
    CORRESPONDENCE ADDRESSES:
      ADDRESSEE: MORRISON & FOERSTER
      STREET: 755 PAGE MILL ROAD
      CITY: Palo Alto
      STATE: CA
      COUNTRY: USA
      ZIP: 94304-1018
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette
      COMPUTER: IBM Compatible
      OPERATING SYSTEM: Windows
      SOFTWARE: FASTSEQ for Windows Version 2.0b
      CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/854,811
        FILING DATE: 12-MAY-1997
        CLASSIFICATION: 435
        PRIOR APPLICATION DATA:
          APPLICATION NUMBER: 08/284,784
          FILING DATE: 02-AUG-1994

```

seq\_name: /cgn2-6/pltdata/2/ina/5B.COMB.seq:US-08-854.811

seq documentation block:  
Sequence 42, Application US/08854811  
Patent No. 5914254

GENERAL INFORMATION:  
APPLICANT: Mascarenhas, Desmond  
APPLICANT: Zhang, Yang  
APPLICANT: Olsson, Pamela S.  
APPLICANT: Olsen, David R.  
APPLICANT: Cohen, Pedro A.  
TITLE OF INVENTION: EXPRESSION OF FUSION POLYPEPTIDES  
TITLE OF INVENTION: TRANSPORTED OUT OF THE CITOPLASM  
TITLE OF INVENTION: SEQUENCES  
NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSO for Windows Version 2.0b

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/854.811  
FILING DATE: 12-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/284,784  
FILING DATE: 02-AUG-1994  
APPLICATION NUMBER: 08/100,744  
FILING DATE: 02-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Buflinger, Nicholas S  
REGISTRATION NUMBER: 39,124  
REFERENCE/DOCKET NUMBER: 22095-20275.21

TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792

INFORMATION FOR SEQ ID NO: 42:

```

SEQUENCE CHANGES:
LENGTH:    915 base pairs
TYPE:      nucleic acid
STRANDEDNESS: single
TOPOLOGY:   linear

US-08-854-811-42

alignment_scores:
Quality:     77.50           Length:       112
Ratio:       1.174          Gaps:        2
Percent Similarity: 58.929   Percent Identity: 19.643

alignment_block:
US-09-484-964-2 x US-08-854-811-42 ..

Align seg 1/1 to: US-08-854-811-42 from: 1 to: 915

1 MetSerAspGlnGluAlaIalysProSerThrGluAsnLeuGlyAaspLysLy 17
:::|||| ::::::|:||::: ||||| |
367 ATCAGCACCTTCACACGCAGAAACAATGCCCGCTTCTTGCGAGGGACCAA 411
17 sgInguylutYrile.....LyslenuyVallelleglygnasps 31

```

```

417 AGCGGCGCAGATATACACTTCAACATGATTTGTCTTCGACC 466
31 erseguillehispheylsvalysmethrthr..... 42
467 GCGGTGCGATGAGATTTGTCAGACATTTGCGTAAACCATACA 516
43 .....Hisleulyslyleulysglusertycy 52
517 TTGGAGTTGAATCTTCGATACATCGACAAAGTTAAGTCGAAATTC 566
52 sglnatgnginglyvalprometasnsrleuarphleupheuglyg 69
567 AGACAGAGAGGATATCCCTCCAGATCAACAAGATTGATCTTGCGGTA 616
69 lharqillealaspasnhrisThrProlysgluleuglymetglu 85
617 AGCAGCTGAGAGCGGTAGACGCTGCTGATTACACATTCAGAGAG 666
86 AspvallilegluvaltyrGlngluInthrlygly 97
667 TCACCTTACATCTTGCTTAAGGCTCCGCGTGT 702

seq_name: /cgn2_6/ptodata/2/lna/3b_COMB.seq:us-08-505-486-92
seq_documentation_block:
; Sequence 92, Application US/08505486
; Patent No. 5955573
; GENERAL INFORMATION:
; APPLICANT: Jesse M. Jaynes
; TITLE OF INVENTION: UBQUITIN-LYTIC PEPTIDE FUSION GENE
; TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROTHWELL, FIGS, ERNST & KURZ
; STREET: 555 Thirteenth Street N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1+
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/505,486
; FILING DATE: 21-JUL-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/279,472
; FILING DATE: 22-JUL-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: WALKER, BARBARA W.
; REGISTRATION NUMBER: 35,400
; REFERENCE/DOCKET NUMBER: 2093-117A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1228
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE STRANDED
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: GENOMIC DNA AND OTHER NUCLEIC ACID
; US-08-505-486-92

alignment_scores:

```

---

```

Quality: 77.00 Length: 68
Ratio: 1.604 Gaps: 0
Percent Similarity: 70.588 Percent Identity: 22.059

alignment_block:
US-09-484-964-2 x US-08-505-486-92 ..
Align seg 1/1 to: US-08-505-486-92 from: 1 to: 1228

34 lIehisphelysValysmethrthrHisleulyslyleulysglu 50
||| .....:
||| .....:
957 ATCACTTACAGGTTGAGTCTCCGACACATGACATTCACACCA 1006
50 rTyrCysGlnArgGlnGlyValPrometasnsrleuarphleuphe 67
1007 GATCCAGACAGAGAGGATTCCTCCAGACGACGCGTTGATTTGG 1056
67 lUglYlNArqilleAlaspasnhrisThrProlysgluleuglymetglu 83
1057 CCGGAAAGCGCTTGAGAGTGTGCTACTCTTCCGACATCAACATCCAG 1106
84 GluGluAspvallilegluvaltyrGlnGluInthrlyglyHisserth 100
1107 AAGGATCAACTCTCCATCTCGTCCGTCGTCCGTGTGTGTGATCCGC 1156
100 rVal 101
1157 TGT 1160

seq_name: /cgn2_6/ptodata/2/lna/6a_COMB.seq:us-08-801-028-92
seq_documentation_block:
; Sequence 92, Application US/08801028
; Patent No. 6018102
; GENERAL INFORMATION:
; APPLICANT: JOAN GARBARINO
; APPLICANT: JESSE M. JAYNES
; APPLICANT: WILLIAM BELKNAP
; TITLE OF INVENTION: UBQUITIN-LYTIC PEPTIDE FUSION GENE CONSTRUCTS, PROTEIN PRO
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STEVEN J. HULPOUIST
; ADDRESS: INTELLECTUAL PROPERTY/TECHNOLOGY LAW
; STREET: 200 PARK DRIVE, SUITE 210
; CITY: P.O. BOX 14329
; CITY: RESEARCH TRIANGLE PARK
; STATE: NORTH CAROLINA
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
; OPERATING SYSTEM: APPLE MACINTOSH
; SOFTWARE: M.S. WORD 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,028
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/279,472
; FILING DATE: JULY 22, 1994
; APPLICATION NUMBER: 08/225,476
; FILING DATE: 04-20-94
; APPLICATION NUMBER: 08/225,476
; FILING DATE: 04-08-94
; APPLICATION NUMBER: 08/039,620
; FILING DATE: 06-04-93
; APPLICATION NUMBER: 08/148,491
; FILING DATE: 11-08-93
; APPLICATION NUMBER: 08/148,889
; FILING DATE: 11-08-93
; ATTORNEY/AGENT INFORMATION:
; NAME: WASSERMAN, FRANK S.
; REGISTRATION NUMBER: 34,273

```



Thu Aug 16 08:48:54 2001

us-09-484-964-2.rni

---

Page 10



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 16:37:34 ; Search time 73.79 Seconds  
(without alignments)  
104.264 Million cell updates/sec

Title: us-09-484-964-2

Perfect score: 525  
Sequence: 1 MSDOEAKPSTEDLGDKKQGE.....MEEDVIEVYOEOTGHSIV 101

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : PIR:68:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	280.5	53.4	91	2 JC5582	SM73 protein - Cae
2	244.5	46.6	117	2 T40313	ubiquitin-like mod
3	233.5	44.5	90	2 T43537	ubiquitin-like pro
4	232	44.2	95	2 JC4760	SM73 protein - hum
5	223.5	42.6	100	2 T04807	SM73 protein homol
6	222	42.3	100	2 T04102	smt3 protein - ric
7	218.5	41.6	101	2 S63999	SM73 protein - yea
8	208	39.6	115	2 T09529	ubiquitin-like pro
9	147	28.0	380	2 T00792	hypothetical prote
10	98	18.7	154	2 S55242	polyubiquitin 2 -
11	98	18.7	156	2 C86439	protein T19E23.13
12	96	18.3	78	2 G86254	hypothetical prote
13	93.5	17.8	456	2 S61158	hypothetical prote
14	88.5	16.9	655	1 S50096	probable splicing
15	85.5	16.3	144	2 A96580	hypothetical prote
16	85.5	16.3	406	2 T39786	hypothetical prote
17	82.5	15.7	551	2 C84549	probable ubiquitin
18	81	15.4	77	1 UONVAC	ubiquitin - Autogr
19	81	15.4	77	2 C72854	viral ubiquitin -
20	81	15.4	77	2 T41781	ubiquitin orf13 -
21	81	15.4	264	2 S62909	ubiquitin precursor
22	81	15.4	381	2 A27806	polyubiquitin 5 (c
23	80	15.2	138	1 UODOR	ubiquitin / riboso
24	80	15.2	150	2 T30390	probable ubiquitin
25	80	15.2	154	1 UODOR7	ubiquitin / riboso
26	80	15.2	228	2 D34080	ubiquitin 18 - s11
27	80	15.2	229	2 B27806	ubiquitin (clone 1
28	80	15.2	380	2 C34080	polyubiquitin 5 (c
29	80	15.2	380	2 B34080	polyubiquitin 5 (c

30	80	15.2	442	2 B71083	hypothetical prote
31	80	15.2	532	2 A34080	polyubiquitin 7 (c
32	79	15.0	76	2 S31653	ubiquitin - Glardi
33	79	15.0	78	2 T39965	ubiquitin-like pro
34	79	15.0	80	2 T28305	ORF MSV144 probabl
35	78.5	15.0	138	2 S62740	ubiquitin precursor
36	76.5	14.6	328	2 T24432	hypothetical prote
37	76	14.5	81	2 JN0710	ubiquitin-like pro
38	76	14.5	128	2 S45304	ubiquitin / riboso
39	76	14.5	177	2 T24466	hypothetical prote
40	76	14.5	274	2 G84602	hypothetical prote
41	75.5	14.4	379	2 S55245	polyubiquitin 5 -
42	75	14.3	129	2 JC5226	ubiquitin / riboso
43	75	14.3	133	2 A49768	polyubiquitin - se
44	75	14.3	149	1 UONCR	ubiquitin / riboso
45	75	14.3	154	2 T46654	ubiquitin/S27a fus

## ALIGNMENTS

```

RESULT 1
JC5582
SM73 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 05-Nov-1999
C:Accession: JC5582; T32928
R:Choudhury, B.K.; Li, S.S.L.
Biochem. Biophys. Res. Commun. 234, 788-791, 1997
A:Title: Identification and characterization of the SM73 cDNA and gene from nematode
A:Reference number: JC5582; MUID:97318858
A:Accession: JC5582
A:Molecule type: DNA
A:Residues: 1-91 <CHO>
A:Cross-references: GB:U94830; MID:92341094; PIDN:AAB67608.1; PID:92341095
R:Wamsley, P.; Kramer, J.
submitted to the EMBL Data Library, January 1998
A:Description: The sequence of C. elegans cosmid K12C11.
A:Reference number: Z21248
A:Accession: T32928
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-91 <WAM>
A:Cross-references: EMBL:AF043701; PIDN:AAB97577.1; GSPDB:GN00019; CESP:K12C11.2
A:Experimental source: strain Bristol N2; clone K12C11
C:Comment: This protein is a suppressor of Mif2 mutation in a centromere protein.
C:Genetics:
A:Gene: smt3; CESP:K12C11.2
A:Map position: 1
A:introns: 22/3; 56/2
C:Superfamily: yeast SM73 protein

Query Match 53.4%; Score 280.5; DB 2; Length 91;
Best local similarity 58.8%; Pred. No. 9.4e-21;
Matches 57; Conservative 15; Mismatches 18; Indels 7; Gaps 2;

QY 1 MSDOEAKPSTEDLGDKKQGEYIKLVIGDSSSEIHKYKMTHTLKKIESYQRCGVPMN 60
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1 MDDDAQ-----AGD--NAEYIKIKYVGDSSNEHRYKYGTSMAKLRKSYADRGVAVN 53

QY 61 SLRFLPEGRIADNHTPKREIGMEEDVIEVYOEOTGG 97
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 54 SLRFLPEGRIADNHTPKREIGMEEDVIEVYOEOTGG 90

RESULT 2
T40313
ubiquitin-like modifier protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000
C:Accession: T40313; T43416
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Leclaire, V.; Galibert, F.

```



16-Jul-1999 #text\_change 20-Jun-2000

C:Date: 16-Jul-1999 #sequence\_16-Jul-1999  
C:Accession: T09529  
R:Labrador, E. The EMBL Data Library, October 1997  
submitted to the EMBL Data Library, October 1997  
A:Description: Isolation and characterization of a cDNA sequence CansMT3 from Cicer arietinum  
A:Reference number: Z16717  
A:Accession: T09529  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-115 <LAB>  
A:Cross-references: EMBL:AJ001901  
A:Experimental source: cultivar Castellana; 5 d old seedlings; etiolated epicotyls  
C:Superfamily: yeast SMT3 protein

Query Match	39.6%	Score 208	DB 2	Length 115	1
Best Local Similarity	44.3%	Pred. No. 1.6e-13			
Matches 43	Conservative % 15	Mismatches 37	Indels 2	Gaps	
1	MSDOEAKPSTEDLGKKGE--YIKLVKIGDSEIHFVKYKTKTLIKESYCORGV	58			
QY	1 MSDOEAKPSTEDLGKKGE--YIKLVKIGDSEIHFVKYKTKTLIKESYCORGV	60			
DB	1 MSGATPMPNPPEDEKPPNDAAHILNKAKGGDNEVFRIKNTOLKRLMNAVCORGSVD				
59	MNSLRFLFEGORIADNHTPKRELGMEEEVIVYEOET	95			
QY	1 MNSLRFLFEGORIADNHTPKRELGMEEEVIVYEOET	97			
DB	1 INSIATFLPDGRRLAAEOTPDLELMDGDDEIDAMLHOT				

RESULT 9  
T000792  
hypothetical protein Atg32760 [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F24L7.10  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 16-Feb-2001  
C|Date: T000792; B84373  
A:Accession: T000792; Lin. X.; Ketchum, R.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K  
Rounsley, S.D.; Lin, X.;  
submitted to the EMBL Data library, February 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.  
A:Reference number: Z14204  
A:Accession: T000792  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1380 <ROU>  
A:Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914697  
A:Experimental source: cultivar Columbia  
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y  
R:Lin. X.; Kaul, S.; Cronin, L.A.; Shen, P.; Vannken, S.E.; Umayam, L.; Tallon,  
M.; Koo, H.; Moffat, K.S.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
euss, D.; Niernann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 407, 761-768, 1999  
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: B84737  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1380 <STO>  
A:Cross-references: GB:AEO02093; NID:g2914697; PID:N:AA04487.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: Atg32760; F24L7.10  
A:Map position: 2  
A:Introns: 55/3; 116/1; 161/3; 237/3; 287/3; 357/3

Query Match	28 0%	Score 147;	DB 2;	Length 380;
Best Local Similarity	39.5%;	Pred. No. 5.6e-07;	Indels 0;	Gaps
Matches	30;	Conservative 15;	Mismatches 31;	
QY	22	IKLTKIGQDSSLEIFKVKMTTHKLKESYRCGRCPVNNLSLRELFBCQKIDNNHPKELG	81	
		:         :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :		
DB	300	VLTKKNNQCGEDLYLIGTHAHLKLMSATCTRNLDYSSFYVYNGREIKARQPPALH	359	
QY	82	MEEDVIEVYQETGG	97	
		:   :		
DB	360	MEEDELICVMYELGGG	375	

360 MEEDEICMVELGGG 375

RESULT 10  
S55242  
polyubiquitin 2 - Arabidopsis thaliana  
N:Alternate names: ubiquitin-like protein  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 10-Oct-1995 #sequence\_revision 7  
C:Accession: S55242  
R:Callis, J.; Carpenter, T.; Sun, C.W.; Vesterstra, R.D.  
Genetics 139, 921-939, 1995  
A:Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins  
A:Reference number: S55242; MUID:95229071  
A:Accession: S55242  
A:Molecule type: DNA  
A:Cross-references: EMBL:L05364  
C:Experimental source: ecotype Columbia  
C:Genetics:  
A:Introns: 31/3; 118/2  
C:Superfamily: polyubiquitin 2; ubiquitin homology  
F:1-76/Product: polyprotein; protein degradation  
F:1-76/Domain: ubiquitin #status Predicted <UB1>  
F:77-132/Product: ubiquitin #status Predicted <UB2>  
F:77-132/Domain: ubiquitin homology <UBH2>

Query Match  
Best Local Similarity 18.7%; Score 98; DB 2; Length 154;  
Matches 21; Conservative 21; Mismatches 35; Indels 2; Gaps 1;  
OY 19 GEYIKLVIGDSSSEIHFKVMTHTLKKLESYCQGVPMNSLRFLFEGQRIADNHTPK 78  
Db 76 GTMKVKTLL--TKREIDIEPTDITRIKERVEKEGIPVQORLIYAGKQADADTKAK 133  
OY 79 ELGNEEDVIEVYQEQGTG 97  
Db 134 DYALGGSVHLVIALRG 152

RESULT 11  
C86439  
Protein T19E23.13 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, A.; Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A66141; MUID:21016719  
A:Accession: C86439  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-156 <STO>  
A:Cross-references: GB:AE005172; NID:96692129; PIDN:AAF24594.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T19E23.13  
A:Map position: 1  
C:Superfamily: polyubiquitin 2; ubiquitin homology

Query Match  
Best Local Similarity 18.7%; Score 98; DB 2; Length 156;  
Matches 21; Conservative 21; Mismatches 35; Indels 2; Gaps 1;

OY 19 GEYIKLVIGDSSSEIHFKVMTHTLKKLESYCQGVPMNSLRFLFEGQRIADNHTPK 78  
Db 76 GTMKVKTLL--TKREIDIEPTDITRIKERVEKEGIPVQORLIYAGKQADADTKAK 133  
OY 79 ELGNEEDVIEVYQEQGTG 97  
Db 134 DYALGGSVHLVIALRG 152

RESULT 12  
G86254  
hypothetical protein [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, A.; Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A66141; MUID:21016719  
A:Accession: G86254  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-78 <STO>  
A:Cross-references: GB:AE005172; NID:93157940; PIDN:AMC17623.1; GSPDB:GN00141  
A:Map position: 1

Query Match  
Best Local Similarity 18.3%; Score 96; DB 2; Length 78;  
Matches 19; Conservative 25; Mismatches 30; Indels 2; Gaps 1;  
OY 22 IKLVIGDSSSEIHFKVMTHTLKKLESYCQGVPMNSLRFLFEGQRIADNHTPK 81  
Db 3 IKVKTLL--KQIDIEPTDITRIKERVEKEGIPVQORLIYAGKQADADTKAK 60  
OY 82 MEEDVIEVYQEQGTG 97  
Db 61 LERGSVHLVIALRG 76

RESULT 13  
S61158  
hypothetical protein YDR363w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D9476.1  
C:Species: Saccharomyces cerevisiae  
C>Date: 23-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 23-Mar-2001  
R:Du, Z.  
Submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of S. cerevisiae cosmid 9476.  
A:Reference number: S61148  
A:Accession: S61158  
A:Molecule type: DNA  
A:Residues: 1-456 <DD2>  
A:Cross-references: EMBL:U28372; NID:9849170; PID:9849171; GSPDB:GN00004; MIPS:YDR363  
A:Gene: MIPS:YDR363w  
A:Map position: 4R

Query Match  
Best Local Similarity 17.8%; Score 93.5; DB 2; Length 456;  
Matches 19; Conservative 24; Mismatches 28; Indels 1; Gaps 1;  
OY 20 EYIKLVIGDSSSEIHFKVMTHTLKKLESYCQGVPMNSLRFLFEGQRIADNHTPK 78

Db 382 EYARIALMGQDNKKIYVHRSPFSKIAEYRIGQKLPQKTRVKLLFDHDELDNMCEIA 441  
 QY 79 ELGMEEDYIEV 90  
 Db 442 DQMEDEMDV 453

## RESULT 14

Probable splicing factor Ccprp21 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: S50096; T33806

R:Spikes, D.A.; Kramer, J.; Bingham, P.M.; van Doren, K.

Nucleic Acids Res. 22, 4510-4519, 1994

A:Title: SWAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a

A:Reference number: S50095; MUID:95061415

A:Accession: S50096

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-655 <SP1>

A:Cross-References: EMBL:U09415; NID:9498845; PIDN:AA82164.1; PID:9498846

R:Latreille, P.; Mamsley, P.

submitted to the EMBL Data Library, November 1998

A:Description: The sequence of C. elegans cosmid w07E6.

A:Reference number: 221414

A:Accession: T33806

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-655 <LAT>

A:Cross-References: EMBL:AF106576; PIDN:AAC78179.1; GSPDB:GN00020; CESP:w07E6.4

A:Experimental source: strain Bristol N2; clone w07E6

C:Genetics:

A:Gene: CESP:w07E6.4

A:Map position: 2

A:Introns: 36/1; 132/2; 185/3; 309/1; 402/2; 464/1; 532/3; 598/3

C:Superfamily: Caenorhabditis elegans probable splicing factor Ccprp21

Query Match 16.9%; Score 88.5; DB 1; Length 655;

Best Local Similarity 24.5%; Pred. No. 0.55;

Matches 26; Conservative 20; Mismatches 43; Indels 17; Gaps 3;

QY 7 KPSTEDLGD-----KQGEYIKLV-----IGDSEIHFKVKKTTLKKIKESY 51

Db 549 RPTED--DLIPEDMLKRVNGAISLNVHLPAPEHGMDGSIQFTIQTAPMSELKQOI 606

QY 52 CORQGVPMNSLRFEGORIDNHTPKELGMEEDYIEVQEQTG 97

Db 607 QDRYGMFVGKQKIMSDGLFVKDNMSAFYNLADRTAITYIQVKERGG 652

## RESULT 15

hypothetical protein F1511.4 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: A96580

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Raul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: A96580

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-144 <STO>  
 A:Cross-References: GB:AF005173; NID:94587536; PIDN:AAD25767.1; GSPDB:GN00141

A:Genetics:

A:Gene: F1511.4

A:Map position: 1

Query Match

Best Local Similarity 16.3%; Score 85.5; DB 2; Length 144;

Matches 23; Conservative 27; Mismatches 27; Indels 41; Gaps 3;

QY 18 QGEYIKLVIGQDS-----EIH-----FKV 38

Db 28 EGKRIKLEV--EDSSNTIDKIHGPTRELVDLSPPTDGAIMMIFIKTLTGRTNYEV 85

QY 39 KMTTLKRLKLSYCORQGVPMNSLRFEGORIDNHTPKELGMEEDYIEVQEQTG 96

Db 86 KGSDFIRLAKAKHEKEGIPVEOQRLIFQGRVLDSDSKAISPYNIKHSSTLHITLHOCG 143

Search completed: August 15, 2001, 17:16:54  
 Job time: 2360 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: August 15, 2001, 17:11:49 ; Search time 57.57 Seconds  
(without alignments)  
60.097 Million cell updates/sec

Title: US-09-484-964-2

Perfect score: 525  
Sequence: 1 MSDOEAKPSTEDIGDKKGE.....MEEDVEYVQPGTGHSTV 101

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 segs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	99.4	101	1	SM33_HUMAN
2	280.5	53.4	91	1	P55853 caenorhabd1
3	244.5	46.6	117	1	SM33_SCHPO
4	239.5	45.6	103	1	SM33_HUMAN
5	232	44.2	95	1	SM32_HUMAN
6	223.5	42.6	104	1	SM33_ARATH
7	222	42.3	100	1	SM33_ORYSA
8	218.5	41.6	101	1	SM33_YEAST
9	81	15.4	77	1	UBIL_NPYAC
10	80	15.2	76	1	UBIL_DICDI
11	76	14.5	76	1	UBIQ_ACACA
12	76	14.5	81	1	NED8_HUMAN
13	75	14.5	81	1	NED8_MOUSE
14	75	14.3	76	1	UBIQ_NEUCR
15	75	14.3	76	1	UBIQ_STRPU
16	74	14.1	76	1	UBIQ_ACEEL
17	74	14.1	76	1	UBIQ_AGLNE
18	74	14.1	76	1	UBIQ_CAEEL
19	74	14.1	76	1	UBIQ_PHYTN
20	73.5	13.9	594	1	YJG9_YEAST
21	73	13.9	76	1	UBIQ_CHLRE
22	73	13.9	76	1	UBIQ_HUMAN
23	73	13.9	76	1	UBIQ_TRYCR
24	73	13.9	76	1	UBIQ_YEAST
25	72.5	13.8	476	1	UBIQ_COPCO
26	72	13.7	76	1	UBIQ_ARATH
27	72	13.7	76	1	UBIQ_TETPY
28	72	13.7	93	1	UBIL_NPYOP
29	71	13.5	76	1	UBIQ_EIMBO
30	71	13.5	76	1	UBIQ_SOYBN
31	70.5	13.4	838	1	OS94_MOUSE
32	70.5	13.4	523	1	KPT2_HUMAN
33	70.5	13.4	523	1	KPT2_RAT

34	70	13.3	76	1	UBIQ_EUPEN	P23324 euplotes eu
35	70	13.3	153	1	UCRP_BOVIN	O02741 bos taurus
36	69	13.1	76	1	UBIQ_LETMA	O05550 leishmania
37	69	13.1	76	1	UBIQ_TRYBA	P15174 trypanosoma
38	69	13.1	352	1	CLCE_PSEAE	O09663 caenorhabd1
39	69	13.1	476	1	YSS4_CAEEL	O07002 homo sapien
40	68.5	13.0	380	1	KPT3_HUMAN	P02161 zalophus ca
41	68	13.0	153	1	MYG_ZALCA	O30847 pseudomona
42	68	13.0	352	1	CLCE_PSEAB	O60739 mus musculu
43	68	13.0	355	1	BAG1_MOUSE	P48510 saccharomyc
44	67	12.8	373	1	DSK2_YEAST	O35569 rattus norv
45	67	12.8	868	1	NRG2_RAT	

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	101 AA.
SM33_HUMAN				
ID	SM33_HUMAN			
AC	Q93068; P55856;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	UBIQUITIN-LIKE PROTEIN SWT3C (UBIQUITIN-HOMOLOGY DOMAIN PROTEIN PIC1)			
DE	(UBIQUITIN-LIKE PROTEIN UBL1) (UBIQUITIN-RELATED PROTEIN SUMO-1)			
DE	(GAP MODIFYING PROTEIN 1) (GMP1) (SENTRIN).			
GN	UBL1 OR SWT33 OR SWT3C.			
OS	Homo sapiens (Human), and Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
CC	NCBI_Taxid=9606, 10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human; TISSUE=Brain;			
RX	MEDLINE=97237059; PubMed=9119407;			
RA	Lapenta V., Chitrazzi P., van der Spek P.J., Pizzuti A.,			
RA	Hanaoka F., Brahe C.;			
RT	"SWT3A, a human homologue of the S. cerevisiae SWT3 gene, maps to			
RT	chromosome 21qter and defines a novel gene family.";			
RL	Genomics 40:362-367(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human; TISSUE=Placenta;			
RX	MEDLINE=96400311; PubMed=8806687;			
RA	Boddy M.N., Howe K., Etkin L.D., Solomon E., Freemont P.S.;			
RT	"Ptc 1, a novel ubiquitin-like protein which interacts with the PML			
RT	component of a multiprotein complex that is disrupted in acute			
RT	promyelocytic leukaemia.";			
RL	Oncogene 13:971-982(1996).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human;			
RX	MEDLINE=96411684; PubMed=8812453;			
RA	Shen Z., Pardington-Putymun P.E., Comeaux J.C., Moyzis R.K.,			
RA	Chen D.J.;			
RT	"UBL1, a human ubiquitin-like protein associating with human			
RT	RAD51/RAD52 proteins.";			
RL	Genomics 36:271-279(1996).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human;			
RX	MEDLINE=97148692; PubMed=9019411;			
RA	Mahajan R., Delphin C., Guan T., Gerace L., Melchior F.;			
RT	"A small ubiquitin-related polypeptide involved in targeting RanGAP1			
RT	to nuclear pore complex protein RanBP2.";			
RL	Cell 88:97-107(1997).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human;			
RX	MEDLINE=97133418; PubMed=8978815;			
RA	Matunis M.J., Coutavas E., Blobel G.;			
RT	"A novel ubiquitin-like modification modulates the partitioning of			

RT the Ran-GTPase-activating protein Rangap1 between the cytosol and the  
 RT nuclear pore complex."  
 RL J. Cell Biol. 135:1457-1470(1996).

# SEQUENCE FROM N.A.

RC SPECIES-Human; TISSUE-Placenta;  
 RX MEDLINE-97064180; PubMed-8906799;  
 RA Okura T., Gong L., Kamitani T., Wada T., Okura I., Wei C.F.,  
 RT Chang H.M., Yeh E.T.H.;  
 RT "Protection against Fas/Apo-1- and tumor necrosis factor-mediated  
 RT cell death by a novel protein, sentrin";  
 RL J. Immunol. 157:4277-4281(1996).

# SEQUENCE FROM N.A.

RC SPECIES-Mouse; STRAIN-ICR;  
 RX MEDLINE-98126440; PubMed-9465300;  
 RA Howe K., Williamson J., Boddy M.N., Sheer D., Freemont P.S.,  
 RT Solomon E.;  
 RT "The ubiquitin-homology gene Plc1: characterization of mouse (Plc1)  
 RT and human (UBI1) genes and pseudogenes";  
 RL Genomics 47:92-100(1998).

# STRUCTURE BY NMR.

RC SPECIES-Human;  
 RX MEDLINE-98319859; PubMed-9654451;  
 RA Bayer P., Arndt A., Metzger S., Mahajan R., Melchior F., Jaenicke R.,  
 RT Becker J.;  
 RT "Structure determination of the small ubiquitin-related modifier  
 RT SUMO-1";  
 RL J. Mol. Biol. 280:275-286(1998).

# -1- FUNCTION: ASSOCIATES WITH RAD51/RAD52. INVOLVED IN TARGETING RANGAP1 TO THE NUCLEAR PORE COMPLEX PROTEIN RANBP2.

# -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X99586; CA67898.1; -  
 DR EMBL: U61397; AAB40388.1; -  
 DR EMBL: U38784; AAC50733.1; -  
 DR EMBL: U67122; AAC50996.1; -  
 DR EMBL: U72722; AAB40390.1; -  
 DR EMBL: U83117; AAB39999.1; -  
 DR EMBL: AF033353; AAC39959.1; -  
 DR PDB: 1A5R; 14-OCT-98.  
 DR MIN: 601912; -  
 DR MGD: MGI:1197010; DB11.  
 DR InterPro: IPR000626; -  
 DR PROSITE: PSS0053; UBIQUITIN\_2; 1.  
 DR 3D-structure.  
 KW DOMAIN 20 101 UBIQUITIN-LIKE.  
 FT SEQUENCE 101 AA; 11557 MW; 89BE97D2D054FB33 CRC64;

Query Match 99.4%; Score 522; DB 1; Length 101;  
 Best Local Similarity 99.0%; Pred. No. 2e-44;  
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDEAKSTEDLDGKKGGEYIKLVIGDSSSEIHFVKMTTHLKLKESYCGROGVPNN 60  
 DB 1 MSDEAKSTEDLDGKKGGEYIKLVIGDSSSEIHFVKMTTHLKLKESYCGROGVPNN 60  
 QY 61 SLRFLFGORLADNHTPKELGMEEDVIEVYQDTGSHSTV 101  
 DB 61 SLRFLFGORLADNHTPKELGMEEDVIEVYQDTGSHSTV 101

RESULT 2

SMT3\_CAEEL  
 ID SMT3\_CAEEL STANDARD; PRT; 91 AA.  
 AC P55853;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE UBIQUITIN-LIKE PROTEIN SMT3.  
 GN SMT3 OR K12C11.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97237059; PubMed-9119407;  
 RA Lapenta V., Chiuazzoli P., van der Spek P.J., Pizzuti A.,  
 RA Hanaoka F., Brahe C.;  
 RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to  
 RT chromosome 21qter and defines a novel gene family";  
 RL Genomics 40:362-367(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97318858; PubMed-9175795;  
 RA Choudhury B.K., Li S.S.;  
 RT "Identification and characterization of the SMT3 cDNA and gene from  
 RT nematode Caenorhabditis elegans";  
 RL Biochem. Biophys. Res. Commun. 234:788-791(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL NZ;  
 RA Wamsley P., Kramer J.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X99600; CA67914.1; -  
 DR EMBL: U94830; AAB67608.1; -  
 DR EMBL: AF043701; AAB97577.1; -  
 DR WormPep: K12C11.2; CE18056.  
 DR InterPro: IPR000626; -  
 DR PROSITE: PSS0053; UBIQUITIN\_2; 1.  
 DR DOMAIN 13 91 UBIQUITIN-LIKE.  
 FT SEQUENCE 91 AA; 10222 MW; 0894E9B6F7B37P5 CRC64;

Query Match 53.4%; Score 280.5; DB 1; Length 91;  
 Best Local Similarity 58.8%; Pred. No. 5.9e-21;  
 Matches 57; Conservative 15; Mismatches 18; Indels 7; Gaps 2;

QY 1 MSDEAKSTEDLDGKKGGEYIKLVIGDSSSEIHFVKMTTHLKLKESYCGROGVPNN 60  
 DB 1 MADDAO-----AGD--NAEYIKLVIGDSSSEIHFVKMTTHLKLKESYCGROGVPNN 53  
 QY 61 SLRFLFGORLADNHTPKELGMEEDVIEVYQDTGSHSTV 97  
 DB 54 SLRFLFGORLADNHTPKELGMEEDVIEVYQDTGSHSTV 90

RESULT 3  
 SMT3\_SCHPO  
 ID SMT3\_SCHPO STANDARD; PRT; 117 AA.  
 AC O13351; 074186;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE UBIQUITIN-LIKE PROTEIN SMT3/PWT3.



```

OS SMT3 OR PMT3 OR SPBC365.06.
OC Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RC MEDLINE=20036821; Pubmed=10567589;
RA Tanaka K., Nishide J., Okazaki K., Kato H., Niwa O., Nakagawa T.,
RA Matsuda H., Kawamukai M., Murakami Y.,
RT "Characterization of a fission yeast SMO-1 homologue, pmt3p,
RT required for multiple nuclear events, including the control of
RT telomere length and chromosome segregation.;"
RL Mol. Cell. Biol. 19:8660-8672(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Jelaire V., Galibert F.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 25-117 FROM N.A.
RC STRAIN=358;
RA Pelletier M.F., Dignard D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR CHROMOSOME SEGREGATION WHERE IT MAY BE
CC INVOLVED IN MICROTUBULE ASSEMBLY. LOSS OF SMT3 LEADS TO AN
CC INCREASE IN TELOMERE LENGTH.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: TO UBIOUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB017187; BAA32595.1; -
DR EMBL; AF019235; AAB71541.1; -
DR EMBL; AL078627; CAB44758.1; ALT_FRAME.
DR InterPro; IPR000626; -
DR PROSITE; PS50053; UBIOUITIN_2; 1.
KW Nuclear protein.
FT DOMAIN 35 115 UBIOUITIN-LIKE.
SQ SEQUENCE 117 AA; 12934 MW; 11C860EBEA172FD2 CRC64;

Query Match 46.6%; Score 244.5; DB 1; Length 117;
Best Local Similarity 52.1%; Pred. No. 2.5e-17;
Matches 50; Conservative 12; Mismatches 25; Indels 9; Gaps 1;

OY 2 SPOEAKPTFEDGDKKQGYIKLYVGOSSSETHFKVMTTLKRLKESYCORQGVPMNS 61
DB 25 SQDDVKKPT-----EHINKYVGGDNNNEVFRIKKTTFESKLMLTYCARQKGSMS 75
OY 62 LRFLEGORIADNHTPKELGMEEDVIEYVQOTG 97
DB 76 LRFIVGERIRPDQTPAELDMEDGDOIIEAVLQDGG 111

RESULT 4
ID SMT3_HUMAN STANDARD; PRT; 103 AA.
AC P55854;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE UBIOUITIN-LIKE PROTEIN SMT3A.
GN SMT3H1 OR SMT3A.

```

```

OS Homo sapiens (Human) .
OC Eukaryota; Metazoa; Chordata; Ciniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP
RP TISSUE=Brain;
RX MEDLINE=97237059; Pubmed=9119407;
RA Lapenta V., Chiurazzi P., van der Spek P.J., Pizzuti A.,
RA Hanacro F., Brahe C.;
RT "SMT3, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL Genomics 40:362-367(1997).
-1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X95984; CAA67896.1; -.
CC MIM; 602231; -.
DR InterPro; IPR000626; -.
DR PROSITE; PS0053; UBIQUITIN_2; 1.
FT DOMAIN 15 UBIQUITIN-LIKE.
SQ SEQUENCE 103 AA; 11654 MW; E335318AC606187 CRC64;

```

```

Query Match Similarity : 45.6%   Score 239.5; DB 1; Length 103;
Best Local Similarity : 47.4%; Pred. No. 6,-6e-17;
Matches 46; Conservative 20; Mismatches 26; Indels 5; Gaps 1.

OY      1 MSDQAKPSTEDLGDKKCGCTIKLKYIGDSSSEIHKVKNMTHTLKTKRSYQQRGVPMN 60
          |:::| - - - - - |:::| 11 11 11 - - - - - |:::| 11 11 11 11 11
DB       1 MSEERKE-----GYVTENDHINKAVAGDSVGVOFKIRKTHSLSKMLTCGERGLSMR 55
OY      61 SLRFEGEGRADNHMPKELGMEEDVIIVYOQTGG 97
          ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB       56 QIRFRDGCPINETDPRAQLRMDEDTIVFQQQTGG 92

RESULT 5
SM32__HUMAN
ID      SM32_HUMAN           STANDARD:             PRT:            95 AA.
AC      P55855;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      UBIQUITIN-LIKE PROTEIN SMT3B (SENTRIN 2).
GN      SMT3H2 OR SMT3B.
OS      Homo sapiens (Human), and Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606, 9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      SPECIES-Human; TISSUE-Brain;
RX      MEDLINE=973737059; PubMed=9119407;
RA      Lapetka V., Chiuwazzl P., van der Spek P.J., Plazut A.,
RA      Hanaka F., Brahe C.;
*SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
chromosome 21qter and defines a novel gene family." ;
RL      Genomics 40:362-367(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      SPECIES-Bovine; TISSUE-Endometrium;
RA      Ol F., Berry E.S.;
RL      Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC      -I- SIMILARITY TO UBIQUITIN, BELONGS TO THE SMT3 SUBFAMILY.

```



```

XX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YPHL/YNN214;
RA Meluh P.B., Koshland D.E.;
RL submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Dietrich F.S.;
RL submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
CC -I- FUNCTION: NOT KNOWN; SUPPRESSOR OF MIF2 MUTATIONS.
CC -I- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U27233; AAB01675.1; .
DR EMBL: U33057; AAB64951.1; .
DR SGD: S0002918; SMT3.
DR InterPro: IPR000626; .
DR PROSITE: PS50053; UBIQUITIN_2; 1.
DR DOMAIN 22 101 UBIQUITIN-LIKE
SQ SEQUENCE 101 AA; 11597 MW; A2790DE7P315E1A7 CRC64;

Query Match      41.6%; Score 218.5; DB 1; Length 101;
Best Local Similarity 48.5%; Pred. No. 7.2e-15;
Matches 49; Conservative 19; Mismatches 26; Indels 7; Gaps 3

QY 1 MSD----QEAKSTEDLDGKOGSEYIKLVIGODSSSEIHKKVMTTLKLKESYCQRQG 56
   |||  ||||| : : : | ||| |||| ||| ||| : : : |||
DQ 1 MSSEVNQGEAKPEVPK--EKYPETHNLKV-SDGSGSEIFFKIKKTPLRLMEAFARQGG 57
   ||| ||||| : : : | ||| |||| ||| ||| : : : |||
DY 57 VPANSLRFLTEFGORIADNHTPKELGMEEBEIVLYVGQCRQG 97
   ||| ||||| : : : | ||| |||| ||| ||| : : : |||
DB 58 KEMDSIRFLTYDGIRIQADDPREDLDMDENDIIEAHRRQIG 98
   ||| ||||| : : : | ||| |||| ||| ||| : : : |||

RESULT 9
UBIL_NPVAC STANDARD; PRT; 77 AA.
AC PI6709;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UBIQUITIN-LIKE PROTEIN.
GN V-UBI.
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolydnavirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90115886; PubMed=2153300;
RA Guarino L.A.;
RL "Identification of a viral gene encoding a ubiquitin-like protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:409-413(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=E2;
RX MEDLINE=86115404; PubMed=3944847;
RA Guarino L.A., Summers M.D.;
RL "Functional mapping of a trans-activating gene required for
RL expression of a baculovirus delayed-early gene.";
RL J Virol. 57:563-571(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
```

```

RA      MDLINE=943031173; PubMed=8030224;
RX      Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT      "the complete DNA sequence of Autographa californica nuclear
RT      polyhedrosis virus.";
RL      Virology 202:586-605(1994).
CC      -1- FUNCTION: UBIQUITIN MAY PLAY A ROLE IN VIRAL LIFE CYCLES, OR
CC      IN VIRUS-HOST INTERACTIONS. IT IS COVALENTLY LINKED TO COAT
CC      PROTEIN SUBUNITS OF SEVERAL DIFFERENT PLANT AND ANIMAL VIRUSES.
CC      -1- SIMILARITY: BELONGS TO THE UBIQUITIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M30305; AAA46751.1; -
DR      EMBL; M37122; AAA46685.1; -
DR      EMBL; L2858; AAA66655.1; -
DR      PIR; A34813; UONVAC.
DR      HSSP; P02248; IAAR.
DR      InterPro; IPR000626; -.
DR      Pfam; PF00240; ubiquitin; 1.
DR      PRINTS; PR00348; UBIQUITIN.
DR      PROSITE; PS00299; UBIQUITIN_1; 1.
DR      PROSITE; PS50053; UBIQUITIN_2; 1.
KW      Nuclear protein; late protein.
FT      SITE 48 48
FT
FT      BINDING 76 76
FT      CONFLICT 20 20
SQ      SEQUENCE 77 AA; 8653 MW;
      EFG9CLID902A8C35A CRC64;
      MULTISUBUNIT ADDUCTS.
      CONFIGURATION TO ACCEPTOR PROTEINS.
      A -> E (IN REF. 1).
      EFG9CLID902A8C35A CRC64;

```

	Query Match	15.4%	Score 81;	DB 1;	Length 77;	
	Best Local Similarity	25.0%;	Pred. No. 0.13;	22;	Indels 0;	Gaps 0
	Matches 13;	Conservative 17;	Mismatches			
Oy	47 LKSEYCORGVPMNMSLRFLEGGRIADNHPTPELMEEEDVIEVYQEOGTGCH 98					
	: :   ::   :  :  :  :  :  :  :					
Dd	26 LKQIAKDREGVPDQORLIIFAGKQLSDSTMDVNIQKESTFLHWLRIRGSGY 77					
	RESULT 10					
	UBIQ_DICDI					
ID	UBIQ_DICDI	STANDARD;	PRT;	76 AA.		
AC	P08618;					
DT	01-AUG-1988 (Rel. 08, Created)					
DT	01-AUG-1988 (Rel. 08, Last sequence update)					
DT	01-NOV-1991 (Rel. 20, Last annotation update)					
DE	UBIQUITIN.					
OS	Dicystostelium discoideum (slime mold).					
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.					
OX	NCBI_TaxID=44689;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=88152253; PubMed=2831095;					
RA	Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,					
RA	Gerisch G.;					
RT	*Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-					
RT	terminal tail and identification of the protein using an anti-peptide					
RL	antibody.";					
FEBS Lett.	229:273-278(1988).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=89352609; PubMed=2548604;					
RA	Ohmachi T., Giorda R., Shaw D.R.; Ennis H.L.;					
RT	"Molecular organization of developmentally regulated Dictyostelium					
RT	discoideum ubiquitin cDNAs ";					
RL	Biochemistry 28:5226-5231(1989).					
RN	[3]					

RP SEQUENCE FROM N.A.  
 RX MEDLINE-87257921; PubMed=3037345;  
 RA Giordano R., Ennis H.L.;  
 RT "Structure of two developmentally regulated Dictyostelium discoideum  
 RL ubiquitin genes.";  
 RM Mol. Cell. Biol. 7:2097-2103(1987).  
 RN [4]  
 RP SEQUENCE OF 13-76 FROM N.A.  
 RC STRAIN-AX2;  
 RA Westphal M., Mueller-Taubenberg A., Noegel A., Gerisch G.;  
 RT "Transcript regulation and carboxyterminal extension of ubiquitin in  
 RL Dictyostelium discoideum.";  
 RM FEBS Lett. 209:32-36(1986).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.  
 CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
 CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES  
 CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X07210; CAA30183.1; ALT\_TERM.  
 DR EMBL: M19666; AAA33261.1; ALT\_TERM.  
 DR EMBL: M19491; AAA33269.1; ALT\_TERM.  
 DR EMBL: M19492; AAA33270.1; ALT\_TERM.  
 DR EMBL: M23748; AAA33262.1; ALT\_TERM.  
 DR EMBL: M23749; AAA33263.1; ALT\_TERM.  
 DR EMBL: M23750; AAA33264.1; ALT\_TERM.  
 DR EMBL: M23751; AAA33265.1; ALT\_TERM.  
 DR EMBL: M23752; AAA33266.1; ALT\_TERM.  
 DR EMBL: M23753; AAA33267.1; ALT\_TERM.  
 DR EMBL: M23754; AAA33268.1; ALT\_TERM.  
 DR EMBL: X04702; CAA28408.1; ALT\_TERM.  
 DR PIR: S00357; UODOR.  
 DR PIR: E34080; UODOR7.  
 DR PIR: A27806; A27806.  
 DR PIR: B27806; B27806.  
 DR PIR: A34080; A34080.  
 DR PIR: B34080; B34080.  
 DR PIR: C34080; C34080.  
 DR HSSP: P02248; 1AAR.  
 DR DICTYDB: DD05001; ubqa.  
 DR InterPro: IPR000626;  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR PROSITE: PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE: PS0053; UBIQUITIN\_2; 1.  
 KW Nuclear protein; Polyprotein.  
 FT SITE 48  
 FT BINDING 76  
 FT VARIATION 11  
 FT SEQUENCE 76 AA; 8538 MW; 6427383968EA84 CRC64;  
 Query Match 15.2%; Score 80; DB 1; Length 76;  
 Best Local Similarity 21.5%; Pred. No. 0.16;  
 Matches 17; Conservative 20; Mismatches 30; Indels 12; Gaps 1;  
 QY 19 GEYIKLAVYIGDSSEIHKVMTHTLKKIKESYCGRGVPMNSLRFLFEQGRADNHTPK 78  
 Db 10 KGTITLEVEGSD-----NINENAKKIDKESIPPDQORLIFAGKOLEGRTLIS 57

QY 79 ELGMEEDVIEVYQETGG 97  
 Db 58 DYNQKRESTHLVLRNG 76  
 RESULT 11  
 ID UBIQ\_ACACA STANDARD; PRT; 76 AA.  
 AC P49634;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE UBIQUITIN.  
 OS Acanthamoeba castellanii (Amoeba).  
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.  
 OX NCBI\_Taxid=5755;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94250685; PubMed=8193157;  
 RA Ahn K.S., Henney H.R.;  
 RT "An Acanthamoeba ubiquitin-fusion protein; cDNA and deduced protein  
 RL sequence.";  
 RT Biochim. Biophys. Acta 1218:109-111(1994).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED WITH RIBOSOMAL CEP52 AS  
 CC ITS C-TERMINAL EXTENSION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X75628; CAA53293.1; ALT\_TERM.  
 DR HSSP: P02248; 1AAR.  
 DR InterPro: IPR000626;  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR PROSITE: PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE: PS0053; UBIQUITIN\_2; 1.  
 KW Nuclear protein; Polyprotein.  
 FT SITE 48  
 FT BINDING 76  
 FT SEQUENCE 76 AA; 8596 MW; D84480E07D1E9B52 CRC64;  
 Query Match 14.5%; Score 76; DB 1; Length 76;  
 Best Local Similarity 20.3%; Pred. No. 0.39;  
 Matches 13; Conservative 21; Mismatches 30; Indels 0; Gaps 0;  
 QY 34 IHFKVMTHTLKKIKESYCGRGVPMNSLRFLFEQGRADNHTPKELGMEEDVIEVYQ 93  
 Db 13 TLEVESDPIENVKQIKDKESIPPDQORLIFAGKOLEGRTLDYNTQKSTHLVLR 72  
 QY 94 QYQG 97  
 Db 73 LRGG 76  
 RESULT 12  
 ID NED8\_HUMAN STANDARD; PRT; 81 AA.  
 AC Q15843;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE UBIOUITIN-LIKE PROTEIN NEDD8.  
 GN NEDD8.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fibrosarcoma;  
 RA Kato S.;  
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=99074275; PubMed=9857030;  
 RA Whitty F.G., Xia G., Pickett C.M., Hill C.P.;  
 RT "Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes."  
 RL J. Biol. Chem. 273:34983-34991(1998).  
 CC -1- FUNCTION: ACTIVATED BY AN E1-LIKE COMPLEX, CONSISTING OF APP-B1 AND UBA3 AND THEN LINKED TO THE E2-LIKE ENZYME, UBC12. THE MAJOR TARGET PROTEIN MODIFIED BY NEDD8 IS CULLIN-4A.  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE DURING THE EMBRYONIC DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM. MAY PLAY AN ESSENTIAL ROLE IN EUKARYOTIC CELLULAR METABOLISM (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE UBIOUITIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D23662; BAA04889.1; -  
 CC PDB; 1NDU; 23-FEB-99.  
 DR MIM; 603171; -  
 DR InterPro; IPR000626; -  
 DR Pfam; PF00240; ubi\_qultin; 1.  
 DR PRINTS; PR00348; UBIOUITIN.  
 DR PROSITE; PS00299; UBIOUITIN\_1; 1.  
 DR PROSITE; PS50053; UBIOUITIN\_2; 1.  
 DR 3D-structure.  
 KW SEQUENCE 81 AA; 9072 MW; DC2FE102BE4725D2 CRC64;  
 SQ  
 Query Match 14.5%; Score 76; DB 1; Length 81;  
 Best Local Similarity 22.4%; Pred. No. 0.42;  
 Matches 17; Conservative 21; Mismatches 36; Indels 2; Gaps 1;  
 QY 22 IKLVIGDSSEIHFVKMTTHLKKIKESYCORGVPMNSLRFLFEGORIADNHTPKELG 81  
 Db 3 IKVYTL-TGKEIIDEIPTDKVERIKEREKEGIPPOORLIYSGKOMDEKTAADYK 60  
 QY 82 MEEDVIEVYQEQTG 97  
 Db 61 ILGGSVLHLVLAIRGG 76  
 Db  
 RESULT 13  
 NEDD8\_MOUSE STANDARD; PRT; 81 AA.  
 AC P29595;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE UBIOUITIN-LIKE PROTEIN NEDD8.  
 GN NEDD8 OR NEDD-8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN

RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=92328780; PubMed=1378265;  
 RA Kumar S., Tomooka Y., Noda M.;  
 RT "Identification of a set of genes with developmentally down-regulated expression in the mouse brain."  
 RL Biochem. Biophys. Res. Commun. 185:1155-1161(1992).  
 RN (2)  
 RP REVISIONS, SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93371434; PubMed=8395831;  
 RA Kumar S., Yoshida Y., Noda M.;  
 RT "Cloning of a cDNA which encodes a novel ubiquitin-like protein."  
 RL Biochem. Biophys. Res. Commun. 195:393-399(1993).  
 CC -1- FUNCTION: ACTIVATED BY AN E1-LIKE COMPLEX, CONSISTING OF APP-B1 AND UBA3 AND THEN LINKED TO THE E2-LIKE ENZYME, UBC12. THE MAJOR TARGET PROTEIN MODIFIED BY NEDD8 IS CULLIN-4A.  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE DURING THE EMBRYONIC DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM. MAY PLAY AN ESSENTIAL ROLE IN EUKARYOTIC CELLULAR METABOLISM.  
 CC -1- TISSUE SPECIFICITY: UBIOUITOUS.  
 CC -1- DEVELOPMENTAL STAGE: DOWN-REGULATED DURING THE DEVELOPMENT OF BRAIN.  
 CC -1- SIMILARITY: BELONGS TO THE UBIOUITIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D10918; BAA01719.1; -  
 CC HSP; P02248; 1AAR.  
 DR MGI; 97301; Nedd8.  
 DR InterPro; IPR000626; -  
 DR Pfam; PF00240; ubi\_qultin; 1.  
 DR PRINTS; PR00348; UBIOUITIN.  
 DR PROSITE; PS00299; UBIOUITIN\_1; 1.  
 DR PROSITE; PS50053; UBIOUITIN\_2; 1.  
 DR SEQUENCE 81 AA; 8972 MW; DC339102BE4725D2 CRC64;  
 SQ  
 Query Match 14.5%; Score 76; DB 1; Length 81;  
 Best Local Similarity 22.4%; Pred. No. 0.42;  
 Matches 17; Conservative 21; Mismatches 36; Indels 2; Gaps 1;  
 QY 22 IKLVIGDSSEIHFVKMTTHLKKIKESYCORGVPMNSLRFLFEGORIADNHTPKELG 81  
 Db 3 IKVYTL-TGKEIIDEIPTDKVERIKEREKEGIPPOORLIYSGKOMDEKTAADYK 60  
 QY 82 MEEDVIEVYQEQTG 97  
 Db 61 ILGGSVLHLVLAIRGG 76  
 Db  
 RESULT 14  
 UBIO\_NEUCR STANDARD; PRT; 76 AA.  
 AC P13117;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE UBIOUITIN.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;  
 RN

```

RX MEDLINE=89366647; PubMed=2549509;
RA Taccioli G.E., Grotewold E., Aisenberg G.O., Judewicz N.D.;
RT "Ubiquitin expression in Neurospora crassa: cloning and sequencing of
RL a polyubiquitin gene.";
RN Nucleic Acids Res. 17:6153-6165(1989).
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;
RX MEDLINE=91323720; PubMed=1650731;
RA Taccioli G.E., Grotewold E., Aisenberg G.O., Judewicz N.D.;
RT "The cDNA sequence and expression of an ubiquitin-tail gene fusion in
RL Neurospora crassa.";
RN Gene 102:133-137(1991).
RM [3]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=94374698; PubMed=8086539;
RA Tarawneh K.A., Amumula K.R., Free S.J.;
RT "The isolation and characterization of a Neurospora crassa gene
RL (ubi::crp-6) encoding a ubiquitin-40S ribosomal protein fusion
protein.";
RM Gene 147:137-140(1994).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH 4 EXACT HEAD TO TAIL REPEATS. THERE IS A FINAL
CC AMINO-ACID (GLN) AFTER THE LAST REPEAT. SOME UBIQUITIN GENES
CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13140; CAA31530.1; ALT_TERM.
DR EMBL: U01220; AAA56880.1; ALT_TERM.
DR EMBL: U01221; AAA03351.1; ALT_TERM.
DR PIR: S05323; UNOC.
DR HSSP: P02248; IAAK.
DR InterPro: IPR000626; -.
DR Pfam: PF00240; ubiquitin; 1.
DR PRINTS: PR00348; UBIQUITIN.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS0053; UBIQUITIN_2; 1.
DR KMW Nuclear protein; Polyprotein.
FT SITE 48
FT 48
FT BINDING 76
FT 76
FT SEQUENCE 76 AA; 8598 MW; 994480FE7D38403E CRC64;
SQ

```

Query Match 14.3%; Score 75; DB 1; Length 76;  
Best Local Similarity 20.3%; Pred. No. 0.49;  
Matches 13; Conservative 20; Mismatches 31; Indels 0; Gaps 0;

```

QY 34 IFKVKMTTLKRLKESYCOQGVPMNSLRFLFEGORIADNHTPEKLGEEEDVLEVOE 93
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 13 IFLVESSDTIDNVKQIKDKGIPPDQORLIFAKQLEDKRTLSDYNIQKSTLHVLVR 72
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 94 QITGG 97
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 73 LRGG 76
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

RESULT 15  
UBIQ\_STRPU  
ID UBIQ\_STRPU STANDARD: PRT; 76 AA.

```

AC P23398;
AT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last annotation update)
DE UBIQUITIN.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinozoa; Echinodermata; Echinozoa; Echinozoa; Echinozoa;
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blasula;
RX MEDLINE=91250006; PubMed=1645680;
RA Nemer M., Rondinelli E., Infante D., Infante A.A.;
RT "Polyubiquitin RNA characteristics and conditional induction in sea
RL urchin embryos.";
RM Dev. Biol. 145:255-265(1991).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. THE NUMBER OF REPEAT
CC DIFFER BETWEEN SPECIES. THERE ARE ESTIMATED TO BE APPROXIMATELY 10
CC AFTER THE LAST REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61172; AAA30082.1; ALT_SEQ.
DR HSSP: P02248; IAAK.
DR InterPro: IPR000626; -.
DR Pfam: PF00240; ubiquitin; 1.
DR PRINTS: PR00348; UBIQUITIN.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS0053; UBIQUITIN_2; 1.
DR KMW Nuclear protein; Polyprotein.
FT SITE 48
FT 48
FT BINDING 76
FT 76
FT SEQUENCE 76 AA; 8551 MW; C42D32397FFD9E02 CRC64;
SQ

```

Query Match 14.3%; Score 75; DB 1; Length 76;  
Best Local Similarity 21.5%; Pred. No. 0.49;  
Matches 17; Conservative 19; Mismatches 31; Indels 12; Gaps 1;

```

QY 19 GEYIKLVYIGDSSSEIFKVKMTTLKRLKESYCOQGVPMNSLRFLFEGORIADNHTPK 78
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 10 GKTITLVEYEDS-----IENVKAKIQDKGIPPDQORLIFAKQLEDKRTLS 57
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 79 ELGMEEDVLEVOETGG 97
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 58 DYNIQKSTLHVLVRNGG 76
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

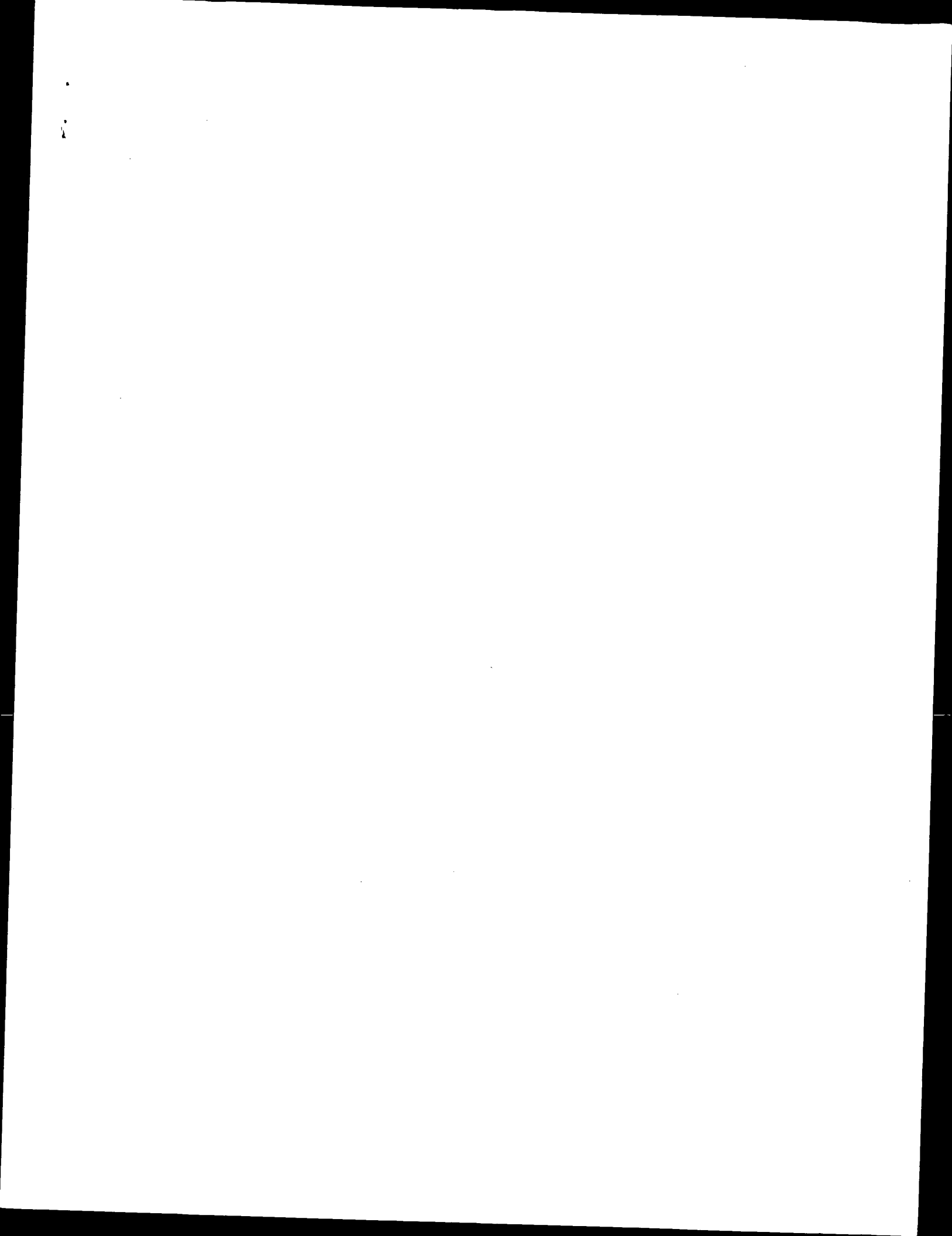
```

Search completed: August 15, 2001, 17:18:03  
Job time: 374 sec

Thu Aug 16 08:48:55 2001

us-09-484-964-2.rsp

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2001, 17:12:59 ; Search time 91.85 Seconds  
(without alignments)  
145.485 Million cell updates/sec

Title: US-09-484-964-2  
Perfect score: 525  
Sequence: 1 MSDDKAPSTEDLDGKKEG.....MEEDVLEVYQEOGTGSHSTV 101

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:REMBL\_16:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.rodent:\*  
13: sp.unclassified:\*  
14: sp.vertebrate:\*  
15: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	517	98.5	101	6	09MZD5
2	470.5	89.6	102	13	057686
3	428	81.5	101	13	09PT08
4	239.5	45.6	110	11	092172
5	235.5	44.9	90	5	097102
6	233	44.4	239	14	057114
7	223.5	42.6	100	10	09S224
8	221	42.1	103	10	09FLP6
9	216.5	41.2	105	10	09SMD1
10	208	39.6	115	10	023759
11	196	37.3	95	4	09S531
12	175.5	33.4	111	10	09FLP5
13	147	28.0	380	10	048841
14	137.5	26.2	114	10	09FKC5
15	120.5	23.0	117	10	09FKC6
16	104	19.8	19	11	09QVNI
17	101	19.2	412	11	009130
18	98.5	18.8	408	6	09GLZ9
19	98	18.7	156	10	09SHE7

20	97.5	18.6	459	11	09JLL1	09j111 rattus norv
21	97.5	18.6	465	11	09JMK4	09jmk4 rattus norv
22	97.5	18.6	465	11	09JMK6	09jmk6 rattus norv
23	96	18.3	78	10	063381	063381 aradidopsis
24	94.5	18.0	255	11	09ES22	09es22 mus musculu
25	94.5	18.0	262	11	09ES23	09es23 mus musculu
26	94.5	18.0	464	11	09WS6	09ws6 mus musculu
27	93.5	17.8	456	3	006340	006340 saccharomyc
28	92.5	17.6	465	4	060260	060260 homo sapien
29	88.5	16.9	655	5	010577	010577 caenorhabdl
30	86	16.4	84	5	09VJ33	09vj33 drosophila
31	86	16.4	154	10	09M551	09m551 populus tre
32	85.5	16.3	144	10	09SYE2	09syf2 aradidopsis
33	85.5	16.3	406	3	09USX3	09usx3 schizosach
34	83	15.8	379	5	027191	027191 tetrahymena
35	83	15.8	902	10	09FH07	09fh07 aradidopsis
36	82.5	15.7	551	10	09S118	09s118 aradidopsis
37	82	15.6	65	5	015871	015871 psalteriomo
38	82	15.6	303	5	027194	027194 tetrahymena
39	81	15.4	80	14	090195	090195 spodoptera
40	81	15.4	264	5	094813	094813 tetrahymena
41	81	15.4	379	5	027192	027192 tetrahymena
42	80	15.2	150	14	09YMT1	09ymt1 lymantiria d
43	80	15.2	349	9	09G044	09g044 bacterioph
44	80	15.2	442	1	058688	058688 pyrococcus
45	79.5	15.1	476	13	013256	013256 gallus gall

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	101 AA.
Q9MZD5	Q9MZD5			
AC	Q9MZD5			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)			
DE	SENTIN.			
OS	Cervus nippon (Sika deer).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae;			
OC	Cervidae; Cervinae; Cervus.			
OX	NCBI_TaxID=9863;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Sun L.G., Yu Y.L., Jiang Y.,			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF242526; AAF97049.1; -			
DR	InterPro: IPR00626; -			
DR	Pfam: PF00240; ubiquitin.1.			
DR	PROSITE: PS50053; Ubiquitin_2.1.			
DR	SEQUENCE 101 AA; 11566 MW; 89A1AAD2D054FB33 CRC64;			
QY	1 MSDDKAPSTEDLDGKKEG...MEEDVLEVYQEOGTGSHSTV 101			
DB	1 MSDDKAPSTEDLDGKKEG...MEEDVLEVYQEOGTGSHSTV 101			
QY	61 SRRFLGGRADNHTPKELGMEEDVLEVYQEOGTGSHSTV 101			
DB	61 SRRFLGGRADNHTPKELGMEEDVLEVYQEOGTGSHSTV 101			
RESULT	2			
ID	057686	PRELIMINARY:	PRT:	102 AA.
AC	057686;			
DT	01-JUN-1998 (TREMBlrel. 06, Created)			







AC 095531;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE D281H8.4 PROTEIN.  
 GN D281H8.4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mashregh-Mohammadi M.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A031133; CA20019.1; -;  
 DR HSSP: Q93068; IASR.  
 DR Interpro: IPR000626; -;  
 DR PROSITE: PS50053; UBIQUITIN\_2; 1.  
 DR SMART: SM00213; UBO; 1.  
 SQ SEQUENCE 95 AA; 10653 MW; 8C7FC29734ED223 CRC64;

Query Match 37.3%; Score 196; DB 4; Length 95;  
 Best Local Similarity 44.0%; Pred. No. 1.3e-11;  
 Matches 40; Conservative 19; Mismatches 30; Indels 2; Gaps 2;

QY 7 KPSTEDGDKKGEYIKLYIGDSSSEIHKVVKMTTHLKLKESYCOROGVPMNSLRLF 66  
 DB 5 KP-EEV-KTENNHNLKVGQDGVQPKIKQTPRLKLMKACPEPRLSVKQIRFRF 62  
 QY 67 EGORIANHNPKEIGMEEDVIEVYQOTGG 97  
 DB 63 GGPISGTDKRPAQLEMEDEDTIVFOQPTGG 93

RESULT 12  
 Q9FLP5 ID Q9FLP5 PRELIMINARY; PRT; 111 AA.  
 AC 09FLP5;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE UBIQUITIN-LIKE PROTEIN.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 NC NCB1\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RX MEDLINE=98290546; PubMed=9628562;  
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen  
 RT physically assigned pl and TAC clones.";  
 RL DNA Res. 5:41-54(1998).  
 DR EMBL: AB010071; BAB08586.1; -;  
 SQ SEQUENCE 111 AA; 12580 MW; 79544A00709B67EC CRC64;

Query Match 33.4%; Score 175.5; DB 10; Length 111;  
 Best Local Similarity 40.4%; Pred. No. 1.3e-09;  
 Matches 38; Conservative 17; Mismatches 34; Indels 5; Gaps 1;

QY 4 QKAPSTEDGDKKGEYIKLYIGDSSSEIHKVVKMTTHLKLKESYCOROGVPMNSLRL 63  
 DB 5 QDDKPI-----DQGEHNVILKAKSODGDEVLFKKNKSAFLKMLATVYCDRSLKIDATA 59  
 QY 64 FLFGORIANHNPKEIGMEEDVIEVYQOTGG 97  
 DB 60 FIFNGARIGLETPDELDMEDGDVIDACRAMSG 93

RESULT 13  
 ID 048841 PRELIMINARY; PRT; 380 AA.  
 AC 048841;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE F24L7.10 PROTEIN.  
 GN F24L7.10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 NC NCB1\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,  
 RA Somerville C.R., Venter J.C.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC003974; AAC04487.1; -;  
 DR HSSP: Q93068; IASR.  
 DR Interpro: IPR000626; -;  
 DR PROSITE: PS50053; UBIQUITIN\_2; 1.  
 DR SMART: SM00213; UBO; 1.  
 SQ SEQUENCE 380 AA; 42844 MW; C8D11F8998517C10 CRC64;

Query Match 28.0%; Score 147; DB 10; Length 380;  
 Best Local Similarity 39.5%; Pred. No. 2.5e-06;  
 Matches 30; Conservative 15; Mismatches 31; Indels 0; Gaps 0;

QY 22 IKLYIGDSSSEIHKVVKMTTHLKLKESYCOROGVPMNSLFLFGORIANHNPKEIG 81  
 DB 300 VILKAVKNOGAGDLYKIGHAHLKLMSATYCKRNLDDYSSVAFYVNGEIKARQTPAQLH 359  
 QY 82 MEEDVIEVYQOTGG 97  
 DB 360 MEEDVIEVYQOTGG 375

RESULT 14  
 Q9FKC5 ID Q9FKC5 PRELIMINARY; PRT; 114 AA.  
 AC 09FKC5;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE SIMILARITY TO SMALL UBIQUITIN-RELATED PROTEIN.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 NC NCB1\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RX MEDLINE=98403884; PubMed=9734815;  
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.  
 RT Sequence features of the regions of 1,367,185 bp covered by 19  
 RT physically assigned pl and TAC clones.";  
 RL DNA Res. 5:203-216(1998).  
 DR EMBL: AB012242; BAB09424.1; -;  
 SQ SEQUENCE 114 AA; 13409 MW; 9D5C8D9154856914 CRC64;

Query Match 26.2%; Score 137.5; DB 10; Length 114;  
 Best Local Similarity 38.5%; Pred. No. 5.6e-06;





2 ATGCTGACCGAGGAGCAAACTTCACTGAGGACTTAGCGATAAGA 51  
 17 sGInGlyuTyriLeuysLeuysValIleGlyGlnAspSerSerGlu 34  
 52 GGAAGGAGATGATTAACCTCAAAAGTATTGGACAGGATGACAGTACA 101  
 34 leHisPheLysValIysMetThrThriHisLeuysLysLeuysGluSer 50  
 102 TACATTTCAAGAGTGAAGATGACAAACATCTCAAGAAACCTCAAGATCA 151  
 51 TyTCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67  
 152 TACGTCAAGAGACAGGAGCTTCCATGAATTCACAGGTTCTCTTGA 201  
 67 uGlyGlnArgGlnLeuAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84  
 202 AGGTGACAGAGATTCGTGATATCATCTCCGAAAGAACTGGGGAATGAGG 251  
 84 lUGluAspValIleGluValTyriGlnGlnGlnThriGlyLysHisSerThr 100  
 252 AAGAGATGATGATTGAAGTTATCAGACAAACAGGGGGTCACTGACG 301  
 101 Val 101  
 302 GTT 304

seq\_name: gb\_est4:AA227354

seq\_documentation\_block:

LOCUS AA227354 361 bp mRNA EST 24-FEB-1997  
 DEFINITION z122612.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens  
 CDNA clone IMAGE:664172 5' similar to TR:G927779 G927779 SUPPRESSOR  
 OF MIF2 MUTATIONS. ; mRNA sequence.

ACCESSION AA227354  
 VERSION AA227354.1 GI:1848909  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 361)  
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldi, M.F., Chapelli, B.,  
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,  
 M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore,  
 B., Morris, M., Parsons, D., Prange, C., Rifkin, L., Rohlfing, T.,  
 Scheinberg, K., Soares, M.B., Tan, F., Thierly-Meg, D., Trevisan, E.,  
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Mair, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu

FEATURES  
 source  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28m1 rev1 ET from Amerisham  
 High quality sequence stop: 282.  
 Location/Qualifiers  
 1..361

/organism="Homo sapiens"  
 /db\_xref="GDB:5426089"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:664172"  
 /clone\_lib="Stratagene NT2 neuronal precursor 937230"  
 /tissue\_type="neuroepithelial cells"  
 /dev\_stage="Ntera-2 neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: Brain; Vector: pBluescript SK-; site\_1:  
 EcoRI; site\_2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. Uninduced, exponentially growing neuroepithelial

BASE COUNT 121 a 68 c 92 g 80 t  
 ORIGIN  
 cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGAG  
 3' ~3' adaptor sequence: 5' CTCAGATTTTTTTTTTTTTTTT 3'

alignment\_scores:  
 Quality: 522.00 Length: 101  
 Ratio: 5.168 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
 US-09-484-964-2 x AA227354

Align seq 1/1 to: AA227354 from: 1 to: 361

1 MetSerAspGlnGlnAlaLysProSerThrGlnAspLeuGlyAspLys 17  
 54 ATGCTGACCGAGGAGCAAACTTCACTGAGGACTTAGCGATAAGA 103  
 17 sGInGlyuTyriLeuysLeuysValIleGlyGlnAspSerSerGlu 34  
 104 GGAAGGAGATGATTAACCTCAAAAGTATTGGACAGGATGACAGTACA 153  
 34 leHisPheLysValIysMetThrThriHisLeuysLysLeuysGluSer 50  
 154 TACATTTCAAGAGTGAAGATGACAAACATCTCAAGAAACCTCAAGATCA 203  
 51 TyTCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67  
 204 TACGTCAAGAGACAGGAGCTTCCATGAATTCACAGGTTCTCTTGA 253  
 67 uGlyGlnArgGlnLeuAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84  
 254 GGGTCAGAGAAATTCGTGATATCATCTCCGAAAGAACTGGGGAATGAGG 303  
 84 lUGluAspValIleGluValTyriGlnGlnGlnThriGlyLysHisSerThr 100  
 304 AAGAGATGATGATTGAAGTTATCAGACAAACAGGGGGTCACTGACG 353  
 101 Val 101  
 354 GTT 356

seq\_name: gb\_est4:AA227355

seq\_documentation\_block:

LOCUS AA227355 368 bp mRNA EST 11-MAR-1998  
 DEFINITION z122612.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens  
 CDNA clone IMAGE:664174 5' similar to TR:G927779 G927779 SUPPRESSOR  
 OF MIF2 MUTATIONS. ; mRNA sequence.

ACCESSION AA227355  
 VERSION AA227355.1 GI:1848910  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 368)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Mair, M., Martin,  
 J., Moore, B., Scheinberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project

TITLE  
 JOURNAL  
 COMMENT  
 Unpublished (1997)  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the



IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert length: 1260 Std Error: 0.00  
 Seq primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 360.

## FEATURES

## Source

1.368  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5426092"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:664174"  
 /clone\_lib="Stratagene NT2 neuronal precursor 937230"  
 /tissue\_type="neuroepithelial cells"  
 /dev\_stage="Ntera-2 neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: brain; Vector: pBluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
 oligo dt. Uninduced, exponentially growing neuroepithelial  
 cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG  
 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 122 a 69 c 91 g 86 t  
 ORIGIN

alignment\_scores:  
 Quality: 522.00 Length: 101  
 Ratio: 5.168 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
 US-09-484-964-2 x AA227355 ..

Align seg 1/1 to: AA227355 from: 1 to: 368

```

1 MetSeraspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
53 ATGCTCACCACGAGCAAAACCTTCACAGAGCTGGGGATTAAGAA 102
17 sGlnGlyGluTyrLleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
103 GGAAGGTGAATATATTAACCTCAAGATGAGACAGATAGAGAGAGA 152
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
153 TTCACCTCAAGTCAAAATGACAAACATCTCAAGAAACCTCAAGATCA 202
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
203 TACTGTCAAAACAGAGGGTGTCCAAATGAATTCACAGGTTCTCTTGA 252
67 uGlyGlnArgLleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
253 GGGTCAGAGAAATGCTGATTAATCACTCCCAAAAGAACTGGGAATGAG 302
84 LuGlnAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThr 100
|||||
303 AAGAAGATGTGATTGAAGTTTATCGAACAACAGGGGGTCAATCAACA 352
101 Val 101
111
353 GTT 355
```

seq\_name: gb\_est17:AI180895

seq\_documentation\_block:

LOCUS AI180895 388 bp mRNA EST 08-OCT-1998  
 DEFINITION ub77a05.t1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 IMAGE:1383728 5' similar to SW:SM33\_HUMAN Q93068 UBIQUITIN-LIKE  
 PROTEIN SWT3C ; mRNA sequence.

ACCESSION AI180895  
 VERSION AI180895.1 GI:3731533  
 KEYWORDS EST.  
 SOURCE house mouse.

## ORGANISM

## REFERENCE

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 388)  
 Mairia, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

## TITLE

The WashU-HMI Mouse EST Project  
 Unpublished (1996)

## JOURNAL

Contact: Mairia M/Mouse EST Project

## COMMENT

WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:906196  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 185.

## FEATURES

## Source

1.368  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1383728"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="MDH10B"  
 /note="Vector: pT73D-pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT73 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 129 a 72 c 103 g 82 t 2 others  
 ORIGIN

alignment\_scores:  
 Quality: 522.00 Length: 101  
 Ratio: 5.168 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
 US-09-484-964-2 x AI180895 ..

Align seg 1/1 to: AI180895 from: 1 to: 388

```

1 MetSeraspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
86 ATGCTCACCACGAGCAAAACCTTCACAGAGCTGGGGATTAAGAA 135
17 sGlnGlyGluTyrLleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
136 GGAAGGTGAATATATTAACCTCAAGATGAGACAGATAGAGAGAGA 185
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
186 TACATTTCAAAGTCAAAATGACAAACATCTCAAGAAACCTCAAGATCA 235
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
236 TACTGTCAAAACAGAGGGTGTCCAAATGAATTCACAGGTTCTCTTGA 285
67 uGlyGlnArgLleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
```

286 AGTCAGAGATTGCTGATATCATCTACCGAAGACTGGCAATGGACG 335  
 84 luguaspvalli1egluval1tyrglglu1nhrcl1ygl1h1s1er1hr 100  
 336 AAGAGAGTGTGATTGAAGTTTATCAGACAAACGGGGGCTCAGTCGACG 385  
 101 Val 101  
 386 GTT 388  
 seq\_name: gb\_est14:AA981271

seq\_documentation\_block:  
 LOCUS AA981271 391 bp mRNA EST 27-MAY-1998  
 DEFINITION vx60f07.r1 Stratagene mouse macrophage (#937306) Mus musculus cDNA  
 clone IMAGE:1279621.5' similar to SW:SM33\_HUMAN Q93068  
 UNROUTIN-LIKE PROTEIN SMT3C ; mRNA sequence.

ACCESSION AA981271  
 VERSION AA981271.1 GI:3159807  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 391)  
 AUTHORS Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 CONTACT: Maria M/Mouse EST Project  
 WASHU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu

THIS CLONE IS AVAILABLE ROYALTY-FREE THROUGH LNL ; CONTACT THE  
 IMAGE CONSORTIUM (info@image.lnl.gov) FOR FURTHER INFORMATION.  
 MGI:671421  
 High quality sequence stop: 379.

## FEATURES

source

Location/Qualifiers

1..391  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_image="1279621"  
 /clone\_id="Stratagene mouse macrophage (#937306)"  
 /tissue\_type="macrophage"  
 /dev\_stage="MEH1-3 cell line"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: blood; Vector: pBluescript SK-; Site: 1:  
 Oligo dT. MEH1-3 cell line. Average insert size: 1.5 kb;  
 Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG  
 3' ~3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"  
 BASE COUNT 128 a 70 c 95 g 97 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 522.00 Length: 101  
 Ratio: 5.168 Gaps: 0  
 Percent Similarity: 100.00 Percent Identity: 99.010

alignment\_block:  
 US-09-484-964-2 x AA981271

Align seg 1/1 to: AA981271 from: 1 to: 391

1 Metserasp1ng1ua1a1y1s1ProserThrg1uasp1euc1y1s1p1s1y1 17

|||||  
 60 ATCTGTGACGAGGAGGCAAAACCTTCACGTGAGGACTTACGCATTAAGAA 109  
 17 sglnglyglu1y1le1y1s1leu1y1val1le1y1gl1n1s1p1s1e1r1g1u1 34  
 110 GCAAGGAGAAATACATTAAACCTCAAAAGTTATTGGACAGATAGCAGTAGA 159  
 34 le1h1s1p1h1e1y1s1v1a1l1y1s1m1e1t1h1r1h1s1leu1y1s1leu1y1g1u1s1e1r1 50  
 160 TACATTTCAAAGTGAATAATGACCAACACATCTCAAGAAACCTCAAGAAATCA 209  
 51 TYTCYSG1n1a1r1g1n1g1y1v1a1Pro1e1t1a1s1e1r1e1u1a1r1g1p1h1e1u1p1e1g1 67  
 210 TACTGTCAAGACGAGGAGGACTTCATTAATGATCACTCAGGTTCTCTTCTTGA 259  
 67 ucl1y1n1a1r1g1l1e1a1a1s1p1a1n1h1s1t1h1r1p1r1o1y1s1g1u1e1u1g1y1m1e1t1g1u1 84  
 260 AGTCAGAGAAATGCTGATATCATCTACCGAAGACTGGCAATGGACG 309  
 84 luguaspvalli1egluval1tyrglglu1nhrcl1ygl1h1s1er1hr 100  
 310 AAGAGAGTGTGATTGAAGTTTATCAGACAAACGGGGGCTCAGTCGACG 359  
 101 Val 101  
 360 GTT 362  
 seq\_name: gb\_est7:AA421208

seq\_documentation\_block:

LOCUS AA421208 402 bp mRNA EST 16-OCT-1997  
 DEFINITION zu05h05.r1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:731001  
 5' similar to TR:G927779 G927779 SUPPRESSOR OF MIF2 MUTATIONS. ;  
 mRNA sequence.  
 ACCESSION AA421208  
 VERSION AA421208.1 GI:2100033  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 402)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marr, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-Merck EST Project 1997  
 Unpublished (1997)  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu

## FEATURES

source

Location/Qualifiers  
 1..402  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5927801"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:731001"  
 /clone\_id="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech Laboratories  
 Inc., and primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATCTGAGTGGAGGCGGCCCATTTTCTTTCTTTCTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT713 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 134 a 74 c 105 g 89 t  
ORIGIN

alignment\_scores: Quality: 522.00 Length: 101  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
US-09-484-964-2 x AA421208 ..

Align seg 1/1 to: AA421208 from: 1 to: 402

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
93 ATGCTGACACGAGGAGGCAAAACCTTCAACTGAGGACTGGGGGTAAGAA 142
17 sGlnGluGluTyrIleLysValIleGlyGlnAspSerSerGlu 34
|||||
143 GGAAGGTGAATATATTAACTCAAAAGTCATTCAGACAGATAGCAGTGAGA 192
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
193 TTCACCTTCAAGTGAAATGACACACATCTCAAGAAACCTCAAGAAATCA 242
51 TyrCysGlnArgGlnGlyValProMetLysSerLeuArgPheLeuPheGlu 67
|||||
243 TACTGTCAAAAGACAGGAGGCTTCCAAATTCACCTGAGTTCTCTTGA 292
67 uGlyGlnArgGluAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
293 GGGTCGACGAATTCGCTGATATCATCTCCAAAGAACTGGGAATGGAGG 342
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyLysHisSerThr 100
|||||
343 AAGAAATGTGATTGAAGTTTATCAGAAACAAACGGGGGTCATTCACACA 392
101 Val 101
111
393 GTT 395

```

seq\_name: gb\_est13:AA896203

seq\_documentation\_block: 426 bp mRNA EST 06-APR-1998  
LOCUS AA896203  
DEFINITION vx62d07.r1 Stragene mouse macrophage (#937306) Mus musculus cDNA  
clone IMAGE:1279789 5' similar to SW:SM33\_HUMAN Q93068  
UBIQUITIN-LIKE PROTEIN SMT3C ; mRNA sequence.

ACCESSION AA896203  
VERSION AA896203.1 GI:3032596  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 426)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:671589  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 407.  
Location/Qualifiers  
1. 426  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1279789"  
/clone\_lib="Stragene mouse macrophage (#937306)"  
/tissue\_type="macrophage"  
/dev\_stage="WEHI-3 cell line"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="organ: blood; Vector: pBluescript SK-; Site: 1:  
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. WEHI-3 cell line. Average insert size: 1.5 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 139 a 80 c 114 g 93 t  
ORIGIN

alignment\_scores: Quality: 522.00 Length: 101  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
US-09-484-964-2 x AA896203 ..

Align seg 1/1 to: AA896203 from: 1 to: 426

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
116 ATGCTGACACGAGGAGGCAAAACCTTCAACTGAGGACTGGGGGTAAGAA 165
17 sGlnGluGluTyrIleLysValIleGlyGlnAspSerSerGlu 34
|||||
166 GGAAGGTGAATATATTAACTCAAAAGTCATTCAGACAGATAGCAGTGAGA 215
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
216 TACATTCAAAAGTGAATATGACACACATCTCAGAAACCTCAAGAAATCA 265
51 TyrCysGlnArgGlnGlyValProMetLysSerLeuArgPheLeuPheGlu 67
|||||
266 TACTGTCAAAAGACAGGAGGCTTCCAAATTCACCTGAGTTCTCTTGA 315
67 uGlyGlnArgGluAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
316 AGGTCAGAGATTGCTGATTAATCATCTCCGAAAGAACTGGGAATGGAGG 365
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyLysHisSerThr 100
|||||
366 AAGAAATGTGATTGAAGTTTATCAGAAACAAACGGGGGTCATTCACAGC 415
101 Val 101
111
416 GTT 418

```

seq\_name: gb\_est15:AI047190

seq\_documentation\_block: 429 bp mRNA EST 08-JUL-1998  
LOCUS AI047190  
DEFINITION uh62d04.r1 Soares-embryonic\_stem\_cell MNES Mus musculus cDNA clone  
IMAGE:1749991 5' similar to SW:SM33\_HUMAN Q93068 UBIQUITIN-LIKE  
PROTEIN SMT3C ; mRNA sequence.

ACCESSION AI047190  
VERSION AI047190.1 GI:3295477  
KEYWORDS EST.





WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:488791  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 422.  
Location/Qualifiers

## FEATURES

## source

1..434  
/organism="Mus musculus"  
/strain="C57BL/6j"  
/db\_xref="taxon:10090"  
/clone="IMAGE:820511"  
/clone\_lib="Soares\_mammary\_gland\_NBMNG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland. Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker. Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dt)  
primer [5'  
TGTTCACATCTGAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M. Fatima  
Bonaldo."

BASE COUNT 138 a 82 c 102 g 112 t  
ORIGIN

## alignment\_scores:

Quality: 522.00 Length: 101  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.010

## alignment block:

US-09-484-964-2 x AA423388 ..

Align seg 1/1 to: AA423388 from: 1 to: 434

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
83 ATGTCTGACGACGAGGCAAAACCTTCACTGAGCACTTAGCGGATTAAGAA 132
17 scInGlyGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerGluI 34
1:::|||||
133 GGAAGGAGATATACATTAACTCAAGTTATTGGACAGATAGCACTGAGA 182
34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
183 TACATTTCAAAGTGAATGACACACATCTCAAGAACTCAAGAAATCA 232
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheGlu 67
|||||
233 TACTGTCAAGACGAGGAGTTCATTAATCATTACACAGGTTCTTCTTGA 282
67 uGlyGlnArgGlyLeuAspAsnHisThrProLysGluLeuGlyMetGluG 84
|||||
283 AGGTGAGGAAATTCGTGATTAATCATATCTCGAAGAACTGGGAATGAGG 332
84 luGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThr 100
|||||
333 AAGAGATGATGATTGAAGTTATATCAGGAACAACGGGGGTCCTCGACG 382
101 Val 101
|||

```

383 GTT 385

seq\_name: gb\_est23:AI646104

seq\_documentation\_block:

LOCUS AI646104 438 bp mRNA EST 29-APR-1999  
DEFINITION U877805.X1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:1383728.3 similar to SW:SM33\_HUMAN Q93068 DB100ITIN-LIKE  
PROTEIN SMT3C; mRNA sequence.

ACCESSION AI646104  
VERSION AI646104.1 GI:4724579  
KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 438)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

TITLE

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:906196

This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 389.

## FEATURES

## source

1..438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1383728"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - oligo(dt) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

Location/Qualifiers

BASE COUNT 127 a 86 c 68 g 157 t  
ORIGIN

## alignment\_scores:

Quality: 522.00 Length: 101  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.010

## alignment block:

US-09-484-964-2 x AI646104/rev ..

Align seg 1/1 to reverse of: AI646104 from: 1 to: 438

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
438 ATGTCTGACGACGAGGCAAAACCTTCACTGAGCACTTAGCGGATTAAGAA 389
17 scInGlyGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerGluI 34
1:::|||||
388 GGAAGGAGATATACATTAACTCAAGTTATTGGACAGATAGCACTGAGA 339
34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
338 TACATTTCAAAGTGAATGACACACATCTCAAGAACTCAAGAAATCA 289

```

51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67  
 |||  
 288 TACTGTCAAGACAGAGGAGTTCACATGATTCATCGAGTTCTCTTGA 239  
 |||  
 67 uG1GlnArg11eG1ValTyrGlnGlnGlnThrGlyGlyHisSerThr 84  
 |||  
 238 AGCTCAGAGAAATTCGTGATATCATCTCCGAAAGAACTGGAGATGAGG 189  
 |||  
 84 1uG1uAspVal11eG1ValTyrGlnGlnGlnThrGlyGlyHisSerThr 100  
 |||  
 188 AAGAGAGATGATTCGATTCATCAGAAACAGGAGGAGTCTGACGAC 139  
 |||  
 101 Val 101  
 |||  
 138 GTT 136

seq\_name: gb\_est86:BF323866

seq\_documentation\_block: 439 bp mRNA EST 21-NOV-2000  
 LOCUS BF323866 SNEST420f08.y1 csn 1 S neuraona invitro mezozoite CDNA Sarcocystis  
 DEFINITION neuraona CDNA 5' similar to SW:SM33\_HUMAN Q93068 UBIQUITIN-LIKE  
 PROTEIN SMT3C ; mRNA sequence.  
 ACCESSION BF323866  
 VERSION BF323866.1 GI:11273469  
 KEYWORDS EST.  
 SOURCE Sarcocystis neuraona.  
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
 Sarcocystis neuraona EST Project.  
 REFERENCE 1 (bases 1 to 439)  
 HOWE,D.K., STAMPER,S., TANG,K., SHIPLEY,L.D., CLIFTON,S., MARRA,M.,  
 HILLIER,L., PAPE,D., MARTIN,J., WYLIE,T., THEISING,B., BOWERS,Y.,  
 GIBBONS,M., RITTER,E., MCCANN,R., BLISTAIN,A., BENNETT,J., SCHMITT,  
 A., KONKO,I., TSAGAREISHVILI,R., FEDELE,M., BELAYGOROD,L.,  
 FRANKLIN,C., CARR,L.M., GROW,A., MAGUIRE,L., WADKINS,J., RICHY,J.,  
 WATERSTON,R. and WILSON,R.  
 Sarcocystis neuraona EST Project  
 Unpublished (2000)  
 CONTACT: Daniel K. Howe  
 Sarcocystis neuraona EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 CONTACT: Daniel K. Howe (dkhowe2@pop.uky.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 376.  
 Location/Qualifiers  
 1..439  
 /organism="Sarcocystis neuraona"  
 /strain="Sn3"  
 /db\_xref="taxon:42890"  
 /clone\_lib="csn 1 S neuraona invitro mezozoite CDNA"  
 /dev\_stage="mezozoite"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI. The library was constructed by Dan Howe, University  
 of Kentucky. cDNAs were synthesized from poly(A)+ RNA  
 by oligo d(T) priming and directionally cloned into the  
 Uni-ZAP XR lambda vector. The library was mass excised  
 as phagemids and rescued in SOLR cells. The plasmid  
 library was recovered from the SOLR cells and transformed  
 in mass into DH10B cells for sequencing. WARNING: the  
 library contains a small percentage of cDNAs derived from  
 the bovine host cells."

BASE COUNT 139 a 81 c 98 g 121 t  
 ORIGIN  
 alignment\_scores:

Quality: 522.00 Length: 101  
 Ratio: 5.168 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment\_block:  
 US-09-484-964-2 x BF323866 ..  
 Align seg 1/1 to: BF323866 from: 1 to: 439

1 MetSerAspGlnGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 17  
 |||  
 62 ATGCTGACACAGAGAGCAAAACCTTCACACGAGACTGGGGGATAGAA 111  
 |||  
 112 GGAAGAGAGTATATTAAACCAAGTCATGACAGAGATGACGTAGA 161  
 |||  
 17 sGlnGlyGlnTyr11eG1ValTyrGlnGlnGlnThrGlyGlyHisSerThr 34  
 |||  
 34 1eHisPheLysVal11eG1ValTyrGlnGlnGlnThrGlyGlyHisSerThr 50  
 |||  
 162 TTCACCTCAAGTGAAGAAATGACACACATCTCAAGAACTCAAGAAATCA 211  
 |||  
 51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67  
 |||  
 212 TACTGTCAAGACAGAGGAGTTCCTCATGATTCACACAGTTCTCTTGA 261  
 |||  
 67 uG1GlnArg11eG1ValTyrGlnGlnGlnThrGlyGlyHisSerThr 84  
 |||  
 262 AGTTCAGAGAAATTCGTGATTCATCAGACATCAAGAAAGTGGAGG 311  
 |||  
 84 1uG1uAspVal11eG1ValTyrGlnGlnGlnThrGlyGlyHisSerThr 100  
 |||  
 312 AAGAAAGATGATTCGATTCATCAGAAACAGGAGGATTCATTCACG 361  
 |||  
 101 Val 101  
 |||  
 362 GTT 364

seq\_name: gb\_est10:AA636685

seq\_documentation\_block: 442 bp mRNA EST 22-OCT-1997  
 LOCUS AA636685 v15f02.r1 Barsstead mouse myotubes MPRB5 Mus musculus CDNA clone  
 DEFINITION IMAGE:1120731 5' similar to TR:G927779 G927779 SUPPRESSOR OF MIR2  
 MUTATIONS. ; mRNA sequence.  
 ACCESSION AA636685  
 VERSION AA636685.1 GI:2560464  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 442)  
 MARRA,M., HILLIER,L., ALLEN,M., BOWLES,M., DIETRICH,N., DUBUQUE,T.,  
 GEISEL,S., KUCABA,T., LACY,M., LE,M., MARTIN,J., MORRIS,M.,  
 SCHEIDENBERG,K., STEPTOE,M., TAN,F., UNDERWOOD,K., MOORE,B.,  
 THEISING,B., WYLIE,T., LENNON,G., SOARES,B., WILSON,R. and  
 WATERSTON,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 CONTACT: Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAG Consortium (info@image.llnl.gov) for further information.  
 MG1:610067  
 Seq primer: -28m3 rev2 ET from Amersham  
 High quality sequence stop: 441.  
 Location/Qualifiers  
 1..442

FEATURES  
 source

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="1120731"
/clone_lib="Barstead mouse myotubes MPRB5"
/cell_line="C2C12"
/lab_host="DH10B"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site 1: EcoRI, Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCGATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pRT3 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."

```

```

BASE COUNT      142 a      82 c      99 g      119 t
ORIGIN
alignment_scores:
  Quality: 522.00      Length: 101
  Ratio: 5.168      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.010

```

```

alignment_block:
US-09-484-964-2 x AA636685

```

```

Align seg 1/1 to: AA636685 from: 1 to: 442

```

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
78 ATGTCGACAGCAGGAGCAAACTTCACCTGAGACTTACGCGATTAAGAA 127
17 sGlnGlyuTyrlleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
128 GGAAGAGAGATTCATTAACTTAAAGTTATGAGACAGATGACAGTGAGA 177
34 lHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
178 TACATTTCAAAGTGAAGATGACACATCTCAGAAACTCAAAAGATCA 227
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
228 TACTGTCAAGACAGGAGGTTCCATGATTCACCTCAGGTTCTCTTGA 277
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
278 AGGTCAGAGAAATTCGTATATCATACCCGAAAGAACTGGGAATGGAG 327
84 lGluAspValIleGluValTyrGlnGlnGlnThrGlyGlnHisSerThr 100
|||||
328 AAGAAGATGTGATTAAGTTATCAGGAACAAACGGGGGTCACTCGACG 377
101 Val 101
||||
378 GTT 380

```

```

seq_name: gb_esc75:BE482568

```

```

seq_documentation_block:

```

```

LOCUS      BE482568      444 bp      mRNA      EST      28-AUG-2000
DEFINITION 168404 BARC 580V Bos taurus cDNA 5', mRNA sequence.
ACCESSION  BE482568
VERSION     BE482568.1 GI:9602101
KEYWORDS   EST.
SOURCE     EST.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bos.

```

```

REFERENCE  1 (bases 1 to 444)
AUTHORS   Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
            Wells,R.D.
TITLE      Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
JOURNAL    Unpublished (2000)
COMMENT    Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bldg. 200, Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8416
            Email: tads@psi.barc.usda.gov
            Single pass sequencing. Bases called and alt-trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 11 row: C column: 20
Seq primer: ATTAGGTGACACTATG.

```

```

FEATURES
  source
  Location/Qualifiers
    1..444
      /organism="Bos taurus"
      /db_xref="taxon:9913"
      /clone_lib="BARC 580V"
      /tissue_type="pooled"
      /lab_host="DH10B"
      /note="Vector: PCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
            library made from pooled mRNA isolated from mammary
            tissues at eight physiological, developmental, and disease
            states."

```

```

BASE COUNT      139 a      86 c      107 g      112 t
ORIGIN

```

```

alignment_scores:
  Quality: 522.00      Length: 101
  Ratio: 5.168      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.010

```

```

alignment_block:
US-09-484-964-2 x BE482568

```

```

Align seg 1/1 to: BE482568 from: 1 to: 444

```

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
96 ATGTCGACAGCAGGAGCAAACTTCACCGAGACTTGGGGATTAAGAA 145
17 sGlnGlyuTyrlleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
146 GGAAGAGAGATATTTAACTCAAGTCATTGGACAGATGACAGTGAAGA 195
34 lHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
196 TTCACTTCAAAAGTGAAGATGACACATCTCAAGAAACTCAAAAGATCA 245
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
246 TACTGTCAAGACAGGAGGTTCCATGATTCACCTCAGGTTCTCTTGA 295
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
296 AGGTCAGAGAAATTCGTATATCATACCTCAGAAAGAACTGGGAATGGAG 345
84 lGluAspValIleGluValTyrGlnGlnGlnThrGlyGlnHisSerThr 100
|||||
346 AAGAAGATGTGATTAAGTTATCAGGAACAAACGGGGGTCACTCAACG 395
101 Val 101
||||
396 GTT 398

```

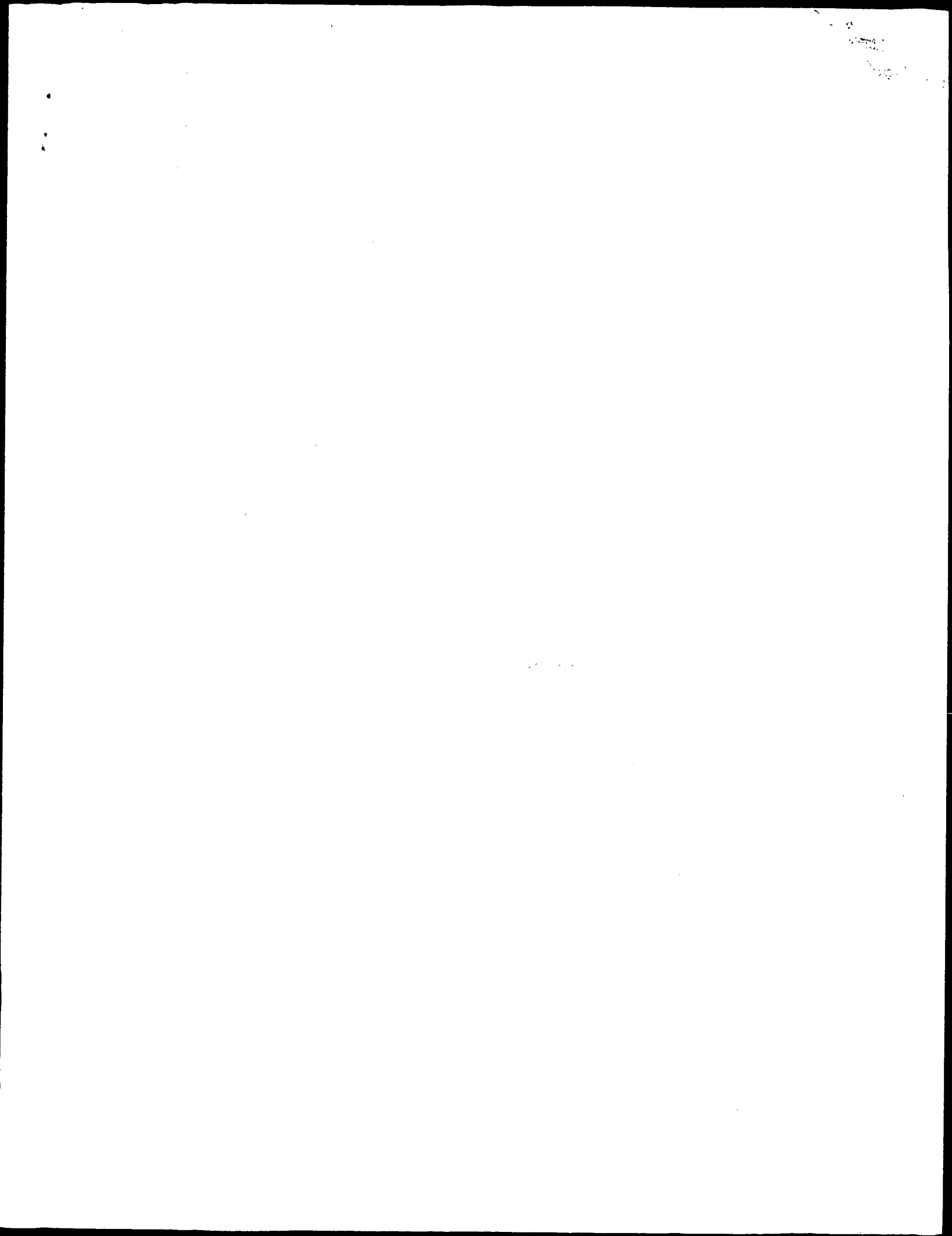


Thu Aug 16 08:48:56 2001

us-09-484-964-2.rst

---

Page 11





REFERENCE 2 (bases 1 to 590)  
 AUTHORS Chintazzi, P.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JUL-1996) P. Chintazzi, Universita' Cattolica - Roma, Largo F. Vito 1, I-00168 Roma, ITALY

FEATURES  
 source Location/Qualifiers  
 1..590  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /dev\_stage="foetus"  
 /tissue\_type="brain"  
 /gene="SMT3C"  
 /gene="SMT3C"  
 /gene="SMT3C"  
 /function="suppressor of MIF2 which encodes a centromere protein"  
 /note="ubiquitin-like protein"  
 /codon\_start=1  
 /product="SMT3C protein"  
 /protein\_id="CAA67898.1"  
 /db\_xref="GI:1770521"  
 /db\_xref="SWISS-PROT:Q93068"  
 /translation="MSDOAKPSTEDLDGKKEGEYIKLVIGDSSEIHFVKMTTHLKKLESYCORQGVPMNSLRFLFGORLADNHHPKEIGMEEDVEIYEOGSHSTV"

BASE COUNT 168 a 113 c 126 g 170 t 13 others

ORIGIN

alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-484-964-2 x HSSMT3C ..

Align seg 1/1 to: HSSMT3C from: 1 to: 590

```

19 G|G|L|U|T|Y|T|L|E|L|Y|S|L|E|U|S|V|A|I|L|E|G|L|G|I|N|A|S|P|S|E|R|G|U|I|E|H|I 35
|||||
96 GGTGAATATATTAATCAAGATCATTTGGACAGATAGCATGATGATCA 145
|||||
35 SPhelysValIysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52
|||||
146 CTTCAGAGTGAAGTGAACACATCTCAAGAACTCAAGAAATCATACT 195
|||||
52 YSGINATG|NG|I|Y|A|I|P|O|M|E|T|A|N|S|E|R|L|E|U|A|T|P|H|E|U|P|H|E|G|I|U|G|I 68
|||||
196 GTCAAGACAGAGGTGTCCAAATGATTCACAGGTTCTCTTGAGGGT 245
|||||
69 G|N|A|T|G|I|E|A|I|A|S|P|A|N|H|I|S|T|H|P|O|L|Y|S|G|I|U|E|U|C|I|Y|E|T|G|I|U|G|I 85
|||||
246 CAGAGAAATTCGTGATATCATATCTCCAAAGAACTGGGAATGGAGGAGA 295
|||||
85 uaspyall|e|g|u|a|l|y|t|g|l|n|g|u|g|l|n|t|p|h|g|l|y|g|l|h|s|e|r|t|p|a|l 101
|||||
296 AGATGTGATTGAAGTTTATCAGGAACAACGGGGGATTCATTCAACAGTT 344
|||||

```

seq\_name: gb\_prl0:HSU67122

seq\_documentation\_block:  
 LOCUS HSU67122 816 bp mRNA PRI 15-FEB-1997  
 DEFINITION Human ubiquitin-related protein SUMO-1 mRNA, complete cds.  
 ACCESSION U67122  
 VERSION U67122.1 GI:1762972  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 816)  
 AUTHORS Mahajan,R., Delphin,C., Guan,T., Gerace,L. and Melchior,F.

TITLE  
 JOURNAL A small ubiquitin-related polypeptide involved in targeting RanGAP1  
 MEDLINE Cell 88 (1), 97-107 (1997)  
 REFERENCE 2 (bases 1 to 816)  
 AUTHORS Mahajan,R.K., Guan,T., Delphin,C., Gerace,L. and Melchior,F.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-AUG-1996) Department of Cell Biology, The Scripps Research Institute, 10666 N. Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES  
 source Location/Qualifiers  
 1..816  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE Consortium Clone ID: 49768, 51818, 199507"

CDS

81..386  
 /note="conjugated post-translationally to RanGAP1; ubiquitin-related protein; similar to UBIL encoded by GenBank Accession Number U61397 and GMP1 encoded by GenBank Accession Number U72722"  
 /codon\_start=1  
 /product="SUMO-1"  
 /protein\_id="AAC50996.1"  
 /db\_xref="GI:1762973"  
 /translation="MSDOAKPSTEDLDGKKEGEYIKLVIGDSSEIHFVKMTTHLKKLESYCORQGVPMNSLRFLFGORLADNHHPKEIGMEEDVEIYEOGSHSTV"

BASE COUNT 235 a 155 c 173 g 255 t

ORIGIN

alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-484-964-2 x HSU67122 ..

Align seg 1/1 to: HSU67122 from: 1 to: 816

```

19 G|G|L|U|T|Y|T|L|E|L|Y|S|L|E|U|S|V|A|I|L|E|G|L|G|I|N|A|S|P|S|E|R|G|U|I|E|H|I 35
|||||
135 GGTGAATATATTAATCAAGATCATTTGGACAGATAGCATGATGATCA 184
|||||
35 SPhelysValIysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52
|||||
185 CTTCAGAGTGAAGTGAACACATCTCAAGAACTCAAGAAATCATACT 234
|||||
52 YSGINATG|NG|I|Y|A|I|P|O|M|E|T|A|N|S|E|R|L|E|U|A|T|P|H|E|U|P|H|E|G|I|U|G|I 68
|||||
235 GTCAAGACAGAGGTGTCCAAATGATTCACAGGTTCTCTTGAGGGT 284
|||||
69 G|N|A|T|G|I|E|A|I|A|S|P|A|N|H|I|S|T|H|P|O|L|Y|S|G|I|U|E|U|C|I|Y|E|T|G|I|U|G|I 85
|||||
285 CAGAGAAATTCGTGATATCATATCTCCAAAGAACTGGGAATGGAGGAGA 334
|||||
85 uaspyall|e|g|u|a|l|y|t|g|l|n|g|u|g|l|n|t|p|h|g|l|y|g|l|h|s|e|r|t|p|a|l 101
|||||
335 AGATGTGATTGAAGTTTATCAGGAACAACGGGGGATTCATTCAACAGTT 383
|||||

```

seq\_name: gb\_prl0:HSU38784

seq\_documentation\_block:  
 LOCUS HSU38784 1017 bp mRNA PRI 16-OCT-1996  
 DEFINITION Human ubiquitin-like protein mRNA, complete cds.  
 ACCESSION U38784  
 VERSION U38784.1 GI:1574947  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1017)  
 AUTHORS Shen, Z., Pardington-Purtymun, P.E., Comeaux, J.C., Moyzis, R.K. and Chen, D.J.  
 TITLE Ubl1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins  
 JOURNAL Genomics 36 (2), 271-276 (1996)  
 MEDLINE 96411684  
 REFERENCE 2 (bases 1 to 1017)  
 AUTHORS Shen, Z.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-OCT-1995) Zhiyuan Shen, Life Sciences Division, Los Alamos National Lab, MS M888, Los Alamos, NM 87545, USA  
 FEATURES  
 source location/Qualifiers  
 1. 1017  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /note="This cDNA was cloned with the yeast two hybrid system. The protein interacts with human RAD51 and RAD52 proteins in yeast"  
 CDS  
 67..372  
 /note="Similar to ubiquitin and to yeast Smtp (suppressor of Mif2); Method: conceptual translation supplied by author"  
 /codon\_start=1  
 /db\_xref="GI:1574948"  
 /translation="MSDEAKPSTEDLGKKEGEYIKLKYIGDSSSEIHFVKMTTHLKKIKESYCOGROGVPNMSLRFLEGORIADNHPKELGMEEDVIEVYQOTGHSIV"  
 BASE COUNT 310 a 180 c 213 g 314 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-484-964-2 x HSU38784 ..  
 Align seg 1/1 to: HSU38784 from: 1 to: 1017

REFERENCE 1 (bases 1 to 1187)  
 AUTHORS Howe, K., Williamson, J., Boddy, N., Sheer, D., Freemont, P. and Solomon, E.  
 TITLE The ubiquitin-homology gene Pici: characterization of mouse (Pici) and human (Ubl1) genes and pseudogenes  
 JOURNAL Genomics 47 (1), 92-100 (1998)  
 MEDLINE 98126440  
 REFERENCE 2 (bases 1 to 1187)  
 AUTHORS Howe, K., Freemont, P.S. and Solomon, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-NOV-1997) Medical & Molecular Genetics, Guy's Hospital, 8th floor, Guy's Tower, London SE19RT, UK  
 FEATURES  
 source location/Qualifiers  
 1. 1187  
 /organism="Mus musculus"  
 /strain="ICR"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /map="C2-3"  
 1..1187  
 /gene="Ubl1"  
 /note="Pici"  
 119..424  
 /gene="Ubl1"  
 /codon\_start=1  
 /product="ubiquitin-homology domain protein"  
 /translation="MSDEAKPSTEDLGKKEGEYIKLKYIGDSSSEIHFVKMTTHLKKIKESYCOGROGVPNMSLRFLEGORIADNHPKELGMEEDVIEVYQOTGHSIV"  
 BASE COUNT 350 a 216 c 267 g 354 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-484-964-2 x AF033353 ..  
 Align seg 1/1 to: AF033353 from: 1 to: 1187



286 GTCAGACACGGGTGTTCCATCATTCACCTACGAGTTCTTTGAGCGT 335  
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85  
 336 CAGAGATTCTCATATCATCTCAAGAAAGACTGGGATGAGAGAGA 385  
 85 uasPValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101  
 386 AGATGTGATTGAAGTTTATCAGAACAAACGGGGGTTCATTCAACAGTT 434

seq\_name: gb\_pat1:AR087051

seq\_documentation\_block:

LOCUS AR087051 1514 bp DNA PAT 07-SEP-2000  
 DEFINITION Sequence 1 from patent US 5985664.  
 ACCESSION AR087051  
 VERSION AR087051.1 GI:10013817

KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1514)

AUTHORS Baker, B.F. and Cowser, L.M.

TITLE Antisense modulation of Sentrin expression

JOURNAL Patent: US 5985664-A 1 16-NOV-1999;  
 FEATURES Location/Qualifiers

source 1..1514

BASE COUNT 462 a 256 c 314 g 482 t  
 ORIGIN /organism="unknown"

alignment\_scores:

Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-484-964-2 x AR087051 ..

Align seg 1/1 to: AR087051 from: 1 to: 1514

19 GtGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH 35  
 190 GGTGAATATATTAACCTCAAGTCATTGGACAGATACAGATGAGATTCA 239  
 35 sPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52  
 240 CTTCAAGTGAATATGACACACATCTCAAGAACTCAAAAGATCATACT 289  
 52 ysgLnaTgGlnGlyValPromeTasnsSerLeuArgPheLeuPheGluGly 68  
 290 GTCAGAACAGAGGTTCTCCAAATGATTCACACAGGTTCTCTTGAGGCT 339  
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85  
 340 CAGAGATTCTCATATCATCTCAAGAAAGACTGGGATGAGAGAGA 389  
 85 uasPValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101  
 390 AGATGTGATTGAAGTTTATCAGAACAAACGGGGGTTCATTCAACAGTT 438

seq\_name: gb\_pat1:HSU083117

seq\_documentation\_block:

LOCUS HSU083117 1514 bp mRNA PRI 09-JAN-1997  
 DEFINITION Human sentrin mRNA, complete cds.  
 ACCESSION U083117  
 VERSION U083117.1 GI:1769601

KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 1514)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS

Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C.-F.,

TITLE

Chang, H.-M., and Yen, E.T.H. Protection Against Fas/Apo-1- and Tumor Necrosis Factor-Mediated

JOURNAL

Cell Death by a Novel Protein, Sentrin

REFERENCE

J. Immunol. 157(10), 4277-4281 (1996)

AUTHORS

Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C.-F.,

TITLE

Chang, H.-M., and Yen, E.T.H. Direct Submission

JOURNAL

Submitted (23-DEC-1996) Division of Molecular Medicine, Department of Internal Medicine, and Cardiovascular Research Center, Institute of Molecular Medicine for the Prevention of Human Diseases, The University of Texas-Houston Health Science Center, 2121 W. Holcombe Blvd., Suite 900, Houston, TX 77030, USA

FEATURES

Location/Qualifiers

source 1..1514

CDS

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="placenta"  
 136..441  
 /note="ubiquitin-like protein"  
 /product="sentrin"  
 /codon\_start=1  
 /protein\_id="AAB39999.1"  
 /db\_xref="GI:1769602"

BASE COUNT 462 a 256 c 314 g 482 t  
 ORIGIN /organism="Homo sapiens"

alignment\_scores:

Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-484-964-2 x HSU083117 ..

Align seg 1/1 to: HSU083117 from: 1 to: 1514

19 GtGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH 35  
 190 GGTGAATATATTAACCTCAAGTCATTGGACAGATACAGATGAGATTCA 239  
 35 sPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52  
 240 CTTCAAGTGAATATGACACACATCTCAAGAACTCAAAAGATCATACT 289  
 52 ysgLnaTgGlnGlyValPromeTasnsSerLeuArgPheLeuPheGluGly 68  
 290 GTCAGAACAGAGGTTCTCCAAATGATTCACACAGGTTCTCTTGAGGCT 339  
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85  
 340 CAGAGATTCTCATATCATCTCAAGAAAGACTGGGATGAGAGAGA 389  
 85 uasPValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101  
 390 AGATGTGATTGAAGTTTATCAGAACAAACGGGGGTTCATTCAACAGTT 438

seq\_name: gb\_cm:AF242526

seq\_documentation\_block:

LOCUS AF242526 510 bp mRNA MAM 03-AUG-2000  
 DEFINITION Cervus nippon sentrin mRNA, complete cds.  
 ACCESSION AF242526  
 VERSION AF242526.1 GI:9664276

KEYWORDS  
 SOURCE sika deer.  
 ORGANISM Cervus nippon

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae;  
 REFERENCE 1 (bases 1 to 510)  
 AUTHORS Sun, L.G., Yu, Y.L. and Jiang, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-MAR-2000) Dept. of Immunology, Norman Bethune  
 University of Medical Sciences, Xin Min Street, Changchun, Jilin  
 130021, People's Republic of China  
 LOCATION/Qualifiers  
 1. 510  
 /organism="Cervus nippon"  
 /db\_xref="taxon:9863"  
 /note="authority: Cervus nippon Temminck"  
 129. 434  
 /codon\_start=1  
 /product="sentrin"  
 /protein\_id="AAP97049.1"  
 /db\_xref="GI:9664277"  
 /translation="MSDQEAKPSTEDLGDKKEGEYIKLVIGDSSSEIHPKVMTHL  
 KKLESYCOHGVPMNSLRFLFEGORLADNHFKELGMEEDYIEVYHQTGHSIV"  
 BASE COUNT 154 a 98 c 121 g 137 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 73.00 Length: 73  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-484-964-2 x AF242526 ..  
 Align seg 1/1 to: AF242526 from: 1 to: 510  
 19 GYGLUTYRILELysLeuLysValIleGIyGlnAspSerSergIuIleH 35  
 |||||||  
 183 GGAGAAATATTAACTCAAGTCATTGGACAGATAGCAGTGAATCA 232  
 35 sPhelysValIysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52  
 |||||||  
 233 CTTCAAAGTGAAGATGACGACACATCTCAAGAACTCAAGAAATCACTACT 282  
 52 YSGlnArGInGIyValPromeTasnsSerLeuArGpHeuPheGluGIy 68  
 |||||||  
 283 GTCAAAGACAGGGAGTTCCTATGAATTCACCTCAGGTTCTTCTTGAAAGT 332  
 69 GlnArGIleAlaAspAsnHisThrProLysGluLeuGIyMetGIuGIu 85  
 |||||||  
 333 CAGAGAAATTCCTGATTAATCACTCAAGAAAGAACTGGGAAATGGAGAAAG 382  
 85 uAspValIleGIuValTyr 91  
 |||||||  
 383 AGATGTGATGAAGTTTAT 401  
 seq\_name: gb\_pat1:AX021844  
 seq\_documentation\_block:  
 LOCUS AX021844 372 bp DNA PAT 07-SEP-2000  
 DEFINITION Sequence 18 from Patent EP0861322.  
 ACCESSION AX021844  
 VERSION AX021844.1 GI:10045087  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 372)  
 AUTHORS O'Neill, R. and Palese, P.  
 TITLE Antiviral compounds that inhibit interaction of host cell proteins  
 JOURNAL and viral proteins required for replication  
 MOUNT SINAI MEDICAL CENTER (US)  
 FEATURES Location/Qualifiers

source 1. 372  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 121 a 72 c 97 g 82 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 67.00 Length: 67  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-484-964-2 x AX021844 ..  
 Align seg 1/1 to: AX021844 from: 1 to: 372  
 19 GYGLUTYRILELysLeuLysValIleGIyGlnAspSerSergIuIleH 35  
 |||||||  
 134 GGTCAAATATTAACTCAAGTCATTGGACAGATAGCAGTGAATCA 183  
 35 sPhelysValIysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52  
 |||||||  
 184 CTTCAAAGTGAAGATGACGACACATCTCAAGAACTCAAGAAATCACTACT 233  
 52 YSGlnArGInGIyValPromeTasnsSerLeuArGpHeuPheGluGIy 68  
 |||||||  
 234 GTCAAAGACAGGGAGTTCCTCAATGAATTCACCTCAGGTTCTTCTTGAGGCT 283  
 69 GlnArGIleAlaAspAsnHisThrProLysGluLeuGIyMetGIuGIu 85  
 |||||||  
 284 CAGAGAAATTCCTGATTAATCACTCAAGAAAGAACTGGGAAATGGAGAAAG 333  
 85 u 85  
 334 A 334  
 seq\_name: gb\_htg17:AC084066  
 seq\_documentation\_block:  
 LOCUS AC084066 235411 bp DNA HTG 12-OCT-2000  
 DEFINITION Mus musculus clone RP23-321D1, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 ACCESSION AC084066  
 VERSION AC084066.1 GI:10799415  
 KEYWORDS HTG; HTGS\_Phasel.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 235411)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Mouse  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 235411)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----  
 Project Information  
 Center Project Name: 2351294  
 Center clone name: RPCR-23\_321D1  
 -----  
 Summary Statistics  
 Consensus quality: 214207 bases at least Q40  
 Consensus quality: 223053 bases at least Q30  
 Consensus quality: 225208 bases at least Q20  
 Estimated insert size: 200000; pulse field gel estimation



Estimated insert size: 232611; sum-of-contrigs estimation  
 Quality coverage: 11.56 in Q20 bases; pulse field gel estimation  
 Quality coverage: 9.94 in Q20 bases; sum-of-contrigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 29 contrigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contrigs are represented as  
 \* runs of 'N', but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1210: contrig of 1210 bp in length
* 1211: gap of unknown length
* 1311: contrig of 1930 bp in length
* 3241: gap of unknown length
* 3341: contrig of 1164 bp in length
* 4505: gap of unknown length
* 4605: contrig of 1568 bp in length
* 6173: gap of unknown length
* 6273: contrig of 1082 bp in length
* 7355: gap of unknown length
* 7455: contrig of 1171 bp in length
* 8625: gap of unknown length
* 8725: contrig of 1389 bp in length
* 10114: gap of unknown length
* 10214: contrig of 1877 bp in length
* 10215: gap of unknown length
* 12092: contrig of 1922 bp in length
* 12192: gap of unknown length
* 14114: contrig of 1073 bp in length
* 14214: gap of unknown length
* 15286: contrig of 2453 bp in length
* 15287: gap of unknown length
* 15386: contrig of 2453 bp in length
* 17839: gap of unknown length
* 17939: contrig of 2327 bp in length
* 17940: gap of unknown length
* 20267: contrig of 1635 bp in length
* 20367: gap of unknown length
* 22002: contrig of 2206 bp in length
* 22102: gap of unknown length
* 24308: contrig of 1998 bp in length
* 24408: gap of unknown length
* 26406: contrig of 3186 bp in length
* 26506: gap of unknown length
* 29691: contrig of 4914 bp in length
* 29791: gap of unknown length
* 29792: contrig of 4914 bp in length
* 34705: gap of unknown length
* 34806: contrig of 4944 bp in length
* 39749: gap of unknown length
* 39750: contrig of 5447 bp in length
* 39850: gap of unknown length
* 45297: contrig of 6080 bp in length
* 45397: gap of unknown length
* 51476: contrig of 7432 bp in length
* 51577: gap of unknown length
* 59008: contrig of 7110 bp in length
* 59109: gap of unknown length
* 59109: contrig of 10460 bp in length
* 66219: gap of unknown length
* 66319: contrig of 21220 bp in length
* 76779: gap of unknown length
* 76879: contrig of 15789 bp in length
* 98099: gap of unknown length
* 98199: contrig of 18028 bp in length
* 113988: gap of unknown length
* 114088: contrig of 160284 bp in length
* 132115: gap of unknown length
* 132116: contrig of 27969 bp in length
* 160185: gap of unknown length
* 160285: contrig of 26414 bp in length
* 186699: gap of unknown length
* 186799: contrig of 48613 bp in length.
  
```

FEATURES  
 source  
 1. 235411  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-321D1"

BASE COUNT 62132 a 54445 c 54225 g 61806 t 2803 others  
 ORIGIN

alignment\_scores:  
 Quality: 65.00 Length: 65  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-484-964-2 x AC084066/rev ..

Align seg 1/1 to reverse of: AC084066 from: 1 to: 235411

19 GYGLTYRILLETYSLEULYSVALILEGLVGLASPSEKRLUJLEH 35  
 |||||  
 113002 TTTCAGAGTGAATGACACACATCTCAAGAACTCAAGATCATCT 112953  
 |||||

35 SPHELVSVALYSWETHRTHRLHSLEULYSLEULYSGLUSERTYRC 52  
 |||||

52 YSGINATGNGLYVALPROMETASNSERLEUARGPHEULPHEGLUGLY 68  
 |||||

112952 GTCAAGACAGGAGGATTCATGATTCACCTCAGGTTCTTGAAGGT 112903  
 |||||

69 GINARGILEALASPANHSIHPROLYSGLUENGLUWETGU 83  
 |||||

112902 CAGAGAAATGCTGATTAATCATACATCCGAAAGACTGGAAATGGAG 112858  
 |||||

seq\_name: gb\_htg22:AL513282

seq\_documentation\_block:  
 LOCUS AL513282 155913 bp DNA HTG 26-FEB-2001  
 DEFINITION Homo sapiens chromosome 1 clone RP11-17965, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 8 unordered pieces.

ACCESSION AL513282  
 VERSION AL513282.5 GI:13162044  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 TITLE Direct Submission  
 AUTHORS Submitted (25-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On Feb 28, 2001 this sequence version replaced gi:13161719.

COMMENT  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: BA17965  
 ----- Summary Statistics  
 Sequencing program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 153015 bases at least Q40  
 Consensus quality: 153938 bases at least Q30  
 Consensus quality: 154565 bases at least Q20  
 Insert size: 155213; sum-of-contrigs  
 Insert size: 164818; 1.0% error; agarose-fp  
 Quality coverage: 6.89x in Q20 bases; sum-of-contrigs quality  
 coverage: 7.27x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contrigs. The true order of the pieces is  
 \* is not known and their order in this sequence record is

```

* Arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 20838: contig of 20838 bp in length
* 20839 20938: gap of 100 bp
* 20939 50589: contig of 29651 bp in length
* 50589 50689: gap of 100 bp
* 50689 81638: contig of 30849 bp in length
* 81639 89682: contig of 8044 bp in length
* 89683 89782: gap of 100 bp
* 89783 108105: contig of 18323 bp in length
* 108106 108205: gap of 100 bp
* 108206 149395: contig of 41190 bp in length
* 149396 149485: gap of 100 bp
* 149486 153075: contig of 3580 bp in length
* 153076 153175: gap of 100 bp
* 153176 155913: contig of 2738 bp in length.
Location/Qualifiers
1. 155913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-17965"
/clone_lib="RPC1-11.1"
1. 20838
/note="assembly-fragment:02743
fragment_chain:1
clone_end:77
vector_side:left"
20939. 50589
/note="assembly-fragment:01236
fragment_chain:1"
50689. 81638
/note="assembly-fragment:00478
fragment_chain:2"
81639. 89682
/note="assembly-fragment:03283
fragment_chain:2"
89783. 108105
/note="assembly-fragment:00942
fragment_chain:3"
108206. 149395
/note="assembly-fragment:02750
fragment_chain:3"
149486. 153075
/note="assembly-fragment:02486
fragment_chain:4"
153176. 155913
/note="assembly-fragment:03083
fragment_chain:4"
BASE COUNT 43830 a 34529 c 35326 g 41527 t 701 others
ORIGIN
alignment_scores:
Quality: 43.00 Length: 43
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-484-964-2 x AL513282/rev ..
Align seg 1/1 to reverse of: AL513282 from: 1 to: 155913
19 GAGTGTATTTTAACTCAAGATGACATCTCAAGAACTCAAGATCATACT 136202
136301 GGTCAATATTTAACTCAAGATGACATCTCAAGATGACATCTCA 136252
35 sphenyVallyMetThrHisLeuLysLysLeuLysGlnSerTyrC 52
|||||

```

```

136251 CTTCAAAGTGAATGACACATCTCAAGAACTCAAGATCATACT 136202
52 YSGTATGAGTGTATTTAACTCAAGATGACATCTCAAGATCATACT 136202
|||||
136201 GTCAAGACAGGCGCTTCATGATTTCA 136173
seq_name: gb_htg21:AL359981
seq_documentation_block:
LOCUS AL359981 167440 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-115N23. *** SEQUENCING IN
PROGRESS ***, 7 unordered pieces.
ACCESSION AL359981
VERSION AL359981.10 GI:12331082
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 167440)
REFERENCE
AUTHORS Pavitt,R.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gi:10186747.
COMMENT
----- Genome Center
Center: Sanger Centre
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba115N23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator; ET-amersham; 16% of reads
Dye-terminator Big Dye; 83% of reads
Consensus quality: 165078 bases at least Q40
Consensus quality: 165122 bases at least Q30
Consensus quality: 165518 bases at least Q20
Insert size: 166840; sum-of-contigs
Insert coverage: 138972; 3.5% error; agarose-tp
Quality coverage: 6.32x in Q20 bases; sum-of-contigs quality
coverage: 8.08x in Q20 bases; agarose-tp
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 17991: contig of 17991 bp in length
* 17992 18091: gap of 100 bp
* 18092 34810: contig of 16719 bp in length
* 34811 34910: gap of 100 bp
* 34911 125180: contig of 90270 bp in length
* 125181 125280: gap of 100 bp
* 125281 129032: contig of 3752 bp in length
* 129033 129132: gap of 100 bp
* 129133 132599: contig of 3467 bp in length
* 132600 132699: gap of 100 bp
* 132700 141603: contig of 8904 bp in length
* 141604 141703: gap of 100 bp
* 141704 167440: contig of 25737 bp in length.
Location/Qualifiers
1. 167440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-115N23"
FEATURES
SOURCE

```

```

misc_feature      /clone_11b="RPC1-11.1"
                  1..17991
                  /note="assembly-fragment:00134
                  fragment_chain:1"
misc_feature      18092..34810
                  /note="assembly-fragment:02116
                  fragment_chain:1"
misc_feature      34911..125180
                  /note="assembly-fragment:03000
                  fragment_chain:1"
misc_feature      125281..129032
                  /note="assembly-fragment:01240
                  fragment_chain:1"
misc_feature      129133..132599
                  /note="assembly-fragment:00856
                  fragment_chain:1"
misc_feature      132700..141603
                  /note="assembly-fragment:00528"
                  141704..167440
misc_feature      /note="assembly-fragment:02701"
                  37675 c 35802 g 45906 t 600 others
BASE COUNT      47457 a 37675 c 35802 g 45906 t 600 others
ORIGIN
alignment_scores:
  Quality: 43.00      Length: 43
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-484-964-2 x AL359981/rev ..
  Align seg 1/1 to reverse of: AL359981 from: 1 to: 167440

19 GYGLGUTYRIIElySLeulysValIIeGLyGlnspserserGIuIIeH1 35
   |||||
154373 GGTGAATATTAACTCAAGTCATTGTGACAGTACAGTACAGATTCA 154324
   |||||
35 SphelysValIysMeTThrThHisLeulysLysLeulysGluSerTYRC 52
   |||||
154323 CTTCAAAAGTGAATAATGACACACATCTCAAGAACTCAAGATCATCT 154274
   |||||
52 YSGInArgInGlyValPromeTAsSer 61
   |||||
154273 GTCAAGACAGAGCGCTTCCCAATGAATCA 154245
   |||||
seq_name: gb-pr5:AL35926

seq_documentation_block:
LOCUS      AL35926 158519 bp      DNA      30-NOV-2000
DEFINITION Human DNA sequence from clone RP11-375F2 on chromosome 1 Contains a
             pseudogene similar to UBI1 (ubiquitin-like 1 (sentrin)), a
             pseudogene similar to ribosomal protein L29, ESTs, STSs and GSSs,
             complete sequence.
ACCESSION  AL35926
VERSION    AL35926.12 GI:9801286
KEYWORDS   HMG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 158519)
            Chapman J
            Direct Submission
            Submitted (03-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            Requests: clonerequest@sanger.ac.uk
            On Aug 14, 2000 this sequence version replaced gi:9621473.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with

```

## FEATURES

```

repeat_region     1..158519
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="1"
                  /clone="RP11-375F2"
                  /clone_11b="RPC1-11.2"
                  /note="MIR repeat: matches 306. .514 of consensus"
                  1465..1514
                  /note="25 copies 2 mer tc 72% conserved"
                  1775..1824
                  /note="MIR repeat: matches 85. .135 of consensus"
                  3415..3734
                  /note="L2 repeat: matches 2373. .2707 of consensus"
                  4294..5519
                  /note="L1MB7 repeat: matches 4938. .6173 of consensus"
                  5535..5640
                  /note="L1MB5 repeat: matches 5322. .5428 of consensus"
                  5652..5945
                  /note="AluSg repeat: matches 12. .309 of consensus"
                  5960..5995
                  /note="9 copies 4 mer aagg 88% conserved"
                  5997..6055
                  /note="L1MB5 repeat: matches 5413. .5471 of consensus"
                  6061..6158
                  /note="AluY repeat: matches 214. .311 of consensus"
                  6160..6855
                  /note="L1MB5 repeat: matches 5460. .6168 of consensus"
                  6919..7101
                  /note="L2 repeat: matches 2206. .2403 of consensus"
                  7099..7228
                  /note="L2 repeat: matches 2620. .2750 of consensus"
                  7828..8069
                  /note="MER6A repeat: matches 2. .235 of consensus"
                  8185..8469
                  /note="AluSx repeat: matches 1. .285 of consensus"
                  8551..8624
                  /note="MIR repeat: matches 81. .153 of consensus"
                  8951..9347
                  /note="match: GSS: Em:B40536"
                  10287..10464
                  /note="MIR repeat: matches 28. .215 of consensus"
                  10588..11321
                  /note="L1PA8 repeat: matches 5371. .6158 of consensus"
                  11337..11671
                  /note="MIR repeat: matches 13. .381 of consensus"

```

only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the NCBI database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/MGP/Ch1>

RP11-375F2 is from the library RPC1-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pBAC3.6  
IMPORTANT: This sequence is not the entire insert of clone RP11-375F2. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-375F2 is at 158519 in this sequence. The true right end of clone RP1-10C16 is at 100 in this sequence.

```

repeat_region 11672..11971
/note="AluY repeat: matches 1..295 of consensus"
misc_feature 11855..12218
/note="unidirectional dGTP only"
repeat_region 11972..12090
/note="MLT1D repeat: matches 381..492 of consensus"
repeat_region 12106..12403
/note="AluX repeat: matches 1..297 of consensus"
repeat_region 12419..13151
/note="L2 repeat: matches 1683..2403 of consensus"
misc_feature 13091..13572
/note="match: GSS: Em:AQ87510"
repeat_region 13164..13253
/note="L2 repeat: matches 2576..2669 of consensus"
repeat_region 13254..13403
/note="75 copies 2 mer at 76% conserved"
repeat_region 13267..13410
/note="24 copies 6 mer tatata 78% conserved"
repeat_region 13654..13742
/note="MLT1E repeat: matches 2..81 of consensus"
repeat_region 14045..14226
/note="MLT1E repeat: matches 385..567 of consensus"
complement(14529..15003)
/note="match: GSS: Em:AQ798857"
misc_feature 14901..15366
/note="match: GSS: Em:AQ221861"
repeat_region 15371..15463
/note="MIR repeat: matches 60..153 of consensus"
repeat_region 16536..16633
/note="LIM4 repeat: matches 3093..3194 of consensus"
repeat_region 16839..17215
/note="LIP1 repeat: matches 5767..6143 of consensus"
repeat_region 17225..17961
/note="LIP3 repeat: matches 5410..6146 of consensus"
repeat_region 17962..19671
/note="LIP7 repeat: matches 4077..5788 of consensus"
repeat_region 19672..20271
/note="LIP7 repeat: matches 3477..4076 of consensus"
repeat_region 20277..20462
/note="LIP4 repeat: matches 3403..3585 of consensus"
repeat_region 20479..21023
/note="LIP8 repeat: matches 1540..1003 of consensus"
repeat_region 21083..22136
/note="LIP8 repeat: matches 977..1014 of consensus"
repeat_region 22104..24249
/note="LIM2 repeat: matches 37..2571 of consensus"
repeat_region 24175..24890
/note="L1 repeat: matches 2118..2843 of consensus"
repeat_region 24896..25532
/note="LIM1 repeat: matches 5508..6163 of consensus"
repeat_region 25865..25900
/note="6 copies 6 mer tgggtg 97% conserved"
repeat_region 25903..26184
/note="AluY repeat: matches 1..282 of consensus"
repeat_region 27110..27647
/note="match: SRS: Em:HS045M12S"
misc_feature 27233..27346
/note="L2 repeat: matches 2549..2668 of consensus"
repeat_region 27347..27437
/note="L2 repeat: matches 2064..2155 of consensus"
repeat_region 27438..27487
/note="LIMC/D repeat: matches 5637..5685 of consensus"
repeat_region 27552..27933
/note="LIMC4 repeat: matches 6621..7006 of consensus"
repeat_region 27982..28315
/note="LIMC4 repeat: matches 7094..7422 of consensus"
repeat_region 28256..28552
/note="LIMC5 repeat: matches 7262..7589 of consensus"
repeat_region 28629..28938
/note="AluY repeat: matches 1..310 of consensus"
repeat_region 29045..29218
/note="LIM1 repeat: matches 5338..5450 of consensus"
repeat_region 30188..30300

```

```

repeat_region 30761..30877
/note="MIR repeat: matches 71..190 of consensus"
repeat_region 31253..31494
/note="L2 repeat: matches 2362..2474 of consensus"
gene 31722..32025
/note="L2 repeat: matches 466..712 of consensus"
CDS 31722..32025
/gene="BA375F2.1"
/gene="BA375F2.1"
/note="BA375F2.1 (similar to UBL1 (ubiquitin-like 1
(sentrin)))
match: cDNAs: Em:U72722
match: ESTs: Em:A119122 Em:A1248769
match: proteins: Tr:O9PT08 Tr:O23759 Sw:P55853 Tr:O92172
Sw:O93068 Sw:P55854 Sw:P55855 Sw:O13351 Sw:O12306
Tr:O97102 Tr:O57686"
/codon_start=1
/pseudo
evidence=not_experimental
complement(32129..32630)
/note="match: GSS: Em:AQ747371"
misc_feature 32225..32619
/note="match: GSS: Em:AQ093192"
complement(32404..32630)
/note="match: GSS: Em:AQ195587"
repeat_region 33174..33222
/note="L2 repeat: matches 2624..2668 of consensus"
repeat_region 33291..33725
/note="MLT1F repeat: matches 94..541 of consensus"
repeat_region 35465..35500

```

```

alignment_scores:
  Quality: 42.00 Length: 42
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-484-964-2 x AL135926

```

```

Align seq 1/1 to: AL135926 from: 1 to: 158519

```

```

23 LysLeuLysValIleGlyGlnAspSerSerGluIleHisPheValIly 39
|||||
31788 AAGTCAAACTGATTGGACAGGATACGACGAGATCTCAAACTGAA 31837
|||||
39 SMeTThrHisLeuLysLeuLysLeuLysGluSerTyTCysGlnArgGlnG 56
|||||
31838 AATGACAAACATCTCAAGAACTCAAGAAATCATCTCAAGGACAGG 31887
|||||
56 LysAlaPrometAsnSerLeuArgPhe 64
|||||
31888 GAGTTCATGAAATTCACCTCAGCATTT 31913

```

OM of: US-09-484-964-2 to: N\_Geneseq\_0601:\* out\_format: pls  
Date: Aug 15, 2001 6:43 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODE=frame+.p2n.model -DEV=xlp  
-O=/cgn2.1/USPRO.spool/US09484964/runtat\_14082001\_111933\_27731/app-query.fasta.1.158  
-DB=N\_Geneseq\_0601 -OFMT=fastap -SUFFIX=oli.rng -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=60.000 -XGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oli.rng  
-TRANS=human4.0.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=quality  
-THR.MIN=20 -ALIGN=15 -MODE=LOCAL -OUTFMT=pls -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09484964 @CGN1\_1\_190 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPHY -WAIT -THREADS=1

## Search information block:

Query: US-09-484-964-2  
Query length: 101  
Database: N\_Geneseq\_0601:\*  
Database sequences: 730101  
Database length: 313950809  
Search time (sec): 128.620000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

## score\_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation	..	1465
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV34564				101.00	1905.04	3.8e-98	1465
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV59722				83.00	1563.51	4.1e-79	1196
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV59557				83.00	1563.37	4.1e-79	1220
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ35859				83.00	1561.83	5.0e-79	1514
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAA16215				71.00	1339.61	1.2e-66	616
/SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:AAZ63339				67.00	1267.00	1.3e-62	372
/SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:AAZ59399				67.00	1267.00	1.3e-62	372
/SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:AAA3182				65.00	1227.94	2.0e-60	425
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAA3182				32.00	600.84	1.7e-25	335

seq\_name: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV34564

seq\_documentation\_block:  
ID AAV34564 standard; DNA: 1465 BP.

XX	AAV34564:	
AC		
XX		
DT	11-SEP-1998 (first entry)	
XX		
DE	Homo sapiens sentrin-1 gene.	
XX		
KW	sentrin-1; protection; tumour necrosis factor; tnfr; apoptosis;	
KW	Fas/Apo-induced; tumour cell death; induction; tumour aggressiveness;	
KW	detection; determination; ss.	
XX		
OS	Homo sapiens.	
XX		
XX		
XX	Key	Location/Qualifiers
XX	FT	88..393
XX	FT	/*tag= a
XX	FT	/product= sentrin-1 polypeptide
XX		
PN	W09820038-AL	
PD	14-MAY-1998.	
XX		
XX	05-NOV-1997;	97WO-US20344.
XX		
PR	05-NOV-1996;	96US-0030302.

XX (TEXTA) UNIV TEXAS SYSTEM.

XX yeh ETH;

XX WPI: 1998-286868/25.

XX P-PSDB: AAW60079.

XX New isolated sentrin polypeptide(s) - which inhibit TNF receptor or  
XX Fas/Apo-induced apoptosis, used to develop products for inducing  
XX cell death in tumours

XX Claim 10: Page 76-77; 120pp: English.

XX The sequence is that encoding the sentrin-1 polypeptide.  
XX Sentrin polypeptides have the ability to protect or guard  
XX cells from tumour necrosis factor (TNF) or Fas/Apo induced  
XX cell death (apoptosis). Inhibitors of the sentrin polypeptides,  
XX e.g. antibodies, can be used for inducing cell death,  
XX particularly in tumours. The products can also be used for  
XX determining the aggressiveness of a tumour and for detection and  
XX isolation of products. The sentrin polypeptide can also be used to  
XX detect a ubiquitin conjugating enzyme polypeptide or PML polypeptide.

XX Sequence 1465 BP; 455 A; 248 C; 292 G; 470 T; 0 other;

## alignment\_scores:

Quality:	101.00	Length:	101
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

## alignment\_block:

US-09-484-964-2 x AAV34564 ..

Align seg 1/1 to: AAV34564 from: 1 to: 1465

1	MetSerAspGlnGluAlaLysProSerThrGlnAspLeuGlyAspLysLys	17
88	ATGCTTGACCGAGGACCAAACTTCAAACTGAGCACTGGGGGATAGAA	137
17	scInGluGluTyrIleLysLeuLysValIleGlyGlnAspSerSerLys	34
138	GCAAGGTGAATATTAATCAAAAGTCATGAGCAGGATGCGGTGGA	187
34	LeHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer	50
188	TTCACCTCAAAAGTGAATGACACACATCTCAAGAACTCAAAAGATCA	237
51	TTCYAGLARGInGInGValPrometAsnSerLeuArgPheLeuPheG	67
238	TACTGTCAAGACAGCGGTGTTCATGAAATTCACACAGTTCCTTGA	287
67	uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu	84
288	GGGTCAAGAAATGCTGATATCATCTCCAAAGACATGGGAATGGGG	337
84	InGluAspValIleGluValTyrGlnGlnGlnThrGlyHisSerThr	100
338	AAGAAGATGTGATTGAATTTATCAGAGAACAAACGGGGGTCATTCAACA	387
101	Val 101	
388	GTT 390	

seq\_name: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV59722

seq\_documentation\_block:

ID AAV59722 standard; DNA: 1196 BP.

XX	AAV59722;	
AC		
XX		
DT	19-JAN-1999 (first entry)	

XX DE Human secreted protein gene 47 clone HOGAV75.  
 XX XX  
 KW Human: secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO9839448-A2.  
 PD 11-SEP-1998.  
 XX  
 PF 06-MAR-1998; 98WO-US044493.  
 XX  
 PR 02-OCT-1997; 97US-0061060.  
 PR 07-MAR-1997; 97US-0038621.  
 PR 07-MAR-1997; 97US-0040161.  
 PR 07-MAR-1997; 97US-0040162.  
 PR 07-MAR-1997; 97US-0040163.  
 PR 07-MAR-1997; 97US-0040163.  
 PR 07-MAR-1997; 97US-0040333.  
 PR 07-MAR-1997; 97US-0040334.  
 PR 07-MAR-1997; 97US-0040626.  
 PR 11-APR-1997; 97US-0043311.  
 PR 11-APR-1997; 97US-0043312.  
 PR 11-APR-1997; 97US-0043313.  
 PR 11-APR-1997; 97US-0043314.  
 PR 11-APR-1997; 97US-0043358.  
 PR 11-APR-1997; 97US-0043569.  
 PR 11-APR-1997; 97US-0043576.  
 PR 11-APR-1997; 97US-0043578.  
 PR 11-APR-1997; 97US-0043580.  
 PR 11-APR-1997; 97US-0043589.  
 PR 11-APR-1997; 97US-0043670.  
 PR 11-APR-1997; 97US-0043671.  
 PR 11-APR-1997; 97US-0043672.  
 PR 11-APR-1997; 97US-0043674.  
 PR 23-MAY-1997; 97US-0047492.  
 PR 23-MAY-1997; 97US-0047500.  
 PR 23-MAY-1997; 97US-0047501.  
 PR 23-MAY-1997; 97US-0047502.  
 PR 23-MAY-1997; 97US-0047503.  
 PR 23-MAY-1997; 97US-0047581.  
 PR 23-MAY-1997; 97US-0047582.  
 PR 23-MAY-1997; 97US-0047583.  
 PR 23-MAY-1997; 97US-0047584.  
 PR 23-MAY-1997; 97US-0047585.  
 PR 23-MAY-1997; 97US-0047586.  
 PR 23-MAY-1997; 97US-0047587.  
 PR 23-MAY-1997; 97US-0047588.  
 PR 23-MAY-1997; 97US-0047589.  
 PR 23-MAY-1997; 97US-0047590.  
 PR 23-MAY-1997; 97US-0047592.  
 PR 23-MAY-1997; 97US-0047593.  
 PR 23-MAY-1997; 97US-0047594.  
 PR 23-MAY-1997; 97US-0047595.  
 PR 23-MAY-1997; 97US-0047596.  
 PR 23-MAY-1997; 97US-0047597.  
 PR 23-MAY-1997; 97US-0047598.  
 PR 23-MAY-1997; 97US-0047599.  
 PR 23-MAY-1997; 97US-0047600.  
 PR 23-MAY-1997; 97US-0047601.  
 PR 23-MAY-1997; 97US-0047612.  
 PR 23-MAY-1997; 97US-0047613.  
 PR 23-MAY-1997; 97US-0047614.  
 PR 23-MAY-1997; 97US-0047615.  
 PR 23-MAY-1997; 97US-0047617.

PR 23-MAY-1997; 97US-0047618.  
 PR 23-MAY-1997; 97US-0047632.  
 PR 23-MAY-1997; 97US-0047633.  
 PR 06-JUN-1997; 97US-0048964.  
 PR 06-JUN-1997; 97US-0048974.  
 PR 13-JUN-1997; 97US-0048974.  
 PR 08-JUL-1997; 97US-0049160.  
 PR 16-JUL-1997; 97US-0051926.  
 PR 18-AUG-1997; 97US-0052874.  
 PR 22-AUG-1997; 97US-0055724.  
 PR 22-AUG-1997; 97US-0056630.  
 PR 22-AUG-1997; 97US-0056631.  
 PR 22-AUG-1997; 97US-0056632.  
 PR 22-AUG-1997; 97US-0056636.  
 PR 22-AUG-1997; 97US-0056637.  
 PR 22-AUG-1997; 97US-0056662.  
 PR 22-AUG-1997; 97US-0056664.  
 PR 22-AUG-1997; 97US-0056664.  
 PR 22-AUG-1997; 97US-0056682.  
 PR 22-AUG-1997; 97US-0056872.  
 PR 22-AUG-1997; 97US-0056874.  
 PR 22-AUG-1997; 97US-0056875.  
 PR 22-AUG-1997; 97US-0056876.  
 PR 22-AUG-1997; 97US-0056877.  
 PR 22-AUG-1997; 97US-0056878.  
 PR 22-AUG-1997; 97US-0056879.  
 PR 22-AUG-1997; 97US-0056880.  
 PR 22-AUG-1997; 97US-0056881.  
 PR 22-AUG-1997; 97US-0056882.  
 PR 22-AUG-1997; 97US-0056884.  
 PR 22-AUG-1997; 97US-0056886.  
 PR 22-AUG-1997; 97US-0056887.  
 PR 22-AUG-1997; 97US-0056888.  
 PR 22-AUG-1997; 97US-0056889.  
 PR 22-AUG-1997; 97US-0056892.  
 PR 22-AUG-1997; 97US-0056893.  
 PR 22-AUG-1997; 97US-0056894.  
 PR 22-AUG-1997; 97US-0056903.  
 PR 22-AUG-1997; 97US-0056908.  
 PR 22-AUG-1997; 97US-0056909.  
 PR 22-AUG-1997; 97US-0056910.  
 PR 22-AUG-1997; 97US-0056911.  
 PR 05-SEP-1997; 97US-0057650.  
 PR 05-SEP-1997; 97US-0057669.  
 PR 05-SEP-1997; 97US-0057761.  
 PR 12-SEP-1997; 97US-0058785.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Edner R, Endress GA;  
 PI Feng P, Ferris AM, Flascher CL, Florence KA, Greene JM, Hu JS;  
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 XX  
 DR WPI: 1998-506364/43.  
 DR P-PSDB: AAW74938.  
 XX  
 PT New isolated human genes and the secreted polypeptide(s) they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS Claim 1: Page 449; 721pp; English.  
 XX  
 CC This sequence represents a nucleic acid molecule designated Gene 47 from  
 CC the human cDNA clone HOGAV75 (deposited as clone ATCC 97857 and ATCC  
 CC 209043) which encodes a secreted human protein. The gene can be used to  
 CC generate fusion proteins by linking to the gene to a human  
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of  
 CC the fused protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 186  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AAV59511 for described uses).

XX Sequence 1196 BP; 360 A; 212 C; 259 G; 364 T; 1 other;

alignment\_scores: Quality: 83.00 Length: 83

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-484-964-2 x AAV59722 ..

Align seg 1/1 to: AAV59722 from: 1 to: 1196

19 GIGIUTYRIELYSLEULYSVAIIIEGLYGLNAPSERSERGIUIIEHI 35  
|||||  
188 GGTGATATATTAACTCAAGTCTTTTGACAGATACAGTACGATTC 237

35 SPHELYSVALYSMETTHRHISLEULYSLEULYSGLUSERTYRC 52  
|||||  
238 GTTCAAGTGAAGAAATGACACACATCTCAGAACTCAAGATCATCT 287

52 YGGLARGINGLYVALPROMETASERIEUARGPHELUAPHEGLY 68  
|||||  
288 GTCAAGACAGGGGTGTCATGATTCACCTAGGTTCTTTAGGGT 337

69 GINATGILEALASPASNHSITHRPROLYSGIULEUGLYMETGLUGLI 85  
|||||  
338 CAGAGATTGCTGATATGATCTCCAAAGAACTGGGATGAGAGAGA 387

85 UASPYALLIEGLUVALTYGINGLUGLINTHRGLYGHISSETHRYAL 101  
|||||  
388 AGATGATGATGAAGTTTATCAGAGAACAAAGGGGGCTTCATCAACGTT 436

seq\_name: /SID1/cgdata/geneseq/geneseq/NA1998.DAT:AAV59557

seq\_documentation\_block:

ID AAV59557 standard; DNA: 1220 BP.

XX AAV59557;

XX 06-JAN-1999 (first entry)

XX Human secreted protein gene 47 clone HOGAV75.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
XX developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
XX immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
XX inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
XX osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
XX endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

XX WO9839448-A2.

XX 11-SEP-1998.

XX 06-MAR-1998: 98WO-US04493.

XX 02-OCT-1997: 97US-0061060.

XX 07-MAR-1997: 97US-0038621.

XX 07-MAR-1997: 97US-0040161.

XX 07-MAR-1997: 97US-0040162.

XX 07-MAR-1997: 97US-0040163.

PR 07-MAR-1997; 97US-0040334.  
PR 07-MAR-1997; 97US-0040335.  
PR 07-MAR-1997; 97US-0040626.  
PR 11-APR-1997; 97US-0043311.  
PR 11-APR-1997; 97US-0043312.  
PR 11-APR-1997; 97US-0043313.  
PR 11-APR-1997; 97US-0043314.  
PR 11-APR-1997; 97US-0043568.  
PR 11-APR-1997; 97US-0043569.  
PR 11-APR-1997; 97US-0043576.  
PR 11-APR-1997; 97US-0043578.  
PR 11-APR-1997; 97US-0043580.  
PR 11-APR-1997; 97US-0043669.  
PR 11-APR-1997; 97US-0043670.  
PR 11-APR-1997; 97US-0043671.  
PR 11-APR-1997; 97US-0043672.  
PR 11-APR-1997; 97US-0043674.  
PR 23-MAY-1997; 97US-0047492.  
PR 23-MAY-1997; 97US-0047500.  
PR 23-MAY-1997; 97US-0047501.  
PR 23-MAY-1997; 97US-0047502.  
PR 23-MAY-1997; 97US-0047503.  
PR 23-MAY-1997; 97US-0047504.  
PR 23-MAY-1997; 97US-0047581.  
PR 23-MAY-1997; 97US-0047582.  
PR 23-MAY-1997; 97US-0047583.  
PR 23-MAY-1997; 97US-0047584.  
PR 23-MAY-1997; 97US-0047585.  
PR 23-MAY-1997; 97US-0047586.  
PR 23-MAY-1997; 97US-0047587.  
PR 23-MAY-1997; 97US-0047588.  
PR 23-MAY-1997; 97US-0047589.  
PR 23-MAY-1997; 97US-0047590.  
PR 23-MAY-1997; 97US-0047592.  
PR 23-MAY-1997; 97US-0047593.  
PR 23-MAY-1997; 97US-0047594.  
PR 23-MAY-1997; 97US-0047595.  
PR 23-MAY-1997; 97US-0047596.  
PR 23-MAY-1997; 97US-0047597.  
PR 23-MAY-1997; 97US-0047598.  
PR 23-MAY-1997; 97US-0047599.  
PR 23-MAY-1997; 97US-0047600.  
PR 23-MAY-1997; 97US-0047601.  
PR 23-MAY-1997; 97US-0047612.  
PR 23-MAY-1997; 97US-0047613.  
PR 23-MAY-1997; 97US-0047614.  
PR 23-MAY-1997; 97US-0047615.  
PR 23-MAY-1997; 97US-0047617.  
PR 23-MAY-1997; 97US-0047618.  
PR 23-MAY-1997; 97US-0047619.  
PR 23-MAY-1997; 97US-0047620.  
PR 23-MAY-1997; 97US-0047621.  
PR 23-MAY-1997; 97US-0047622.  
PR 23-MAY-1997; 97US-0047623.  
PR 23-MAY-1997; 97US-0047624.  
PR 23-MAY-1997; 97US-0047625.  
PR 23-MAY-1997; 97US-0047626.  
PR 23-MAY-1997; 97US-0047627.  
PR 23-MAY-1997; 97US-0047628.  
PR 23-MAY-1997; 97US-0047629.  
PR 23-MAY-1997; 97US-0047630.  
PR 23-MAY-1997; 97US-0047631.  
PR 23-MAY-1997; 97US-0047632.  
PR 23-MAY-1997; 97US-0047633.  
PR 23-MAY-1997; 97US-0047634.  
PR 23-MAY-1997; 97US-0047635.  
PR 23-MAY-1997; 97US-0047636.  
PR 23-MAY-1997; 97US-0047637.  
PR 23-MAY-1997; 97US-0047638.  
PR 23-MAY-1997; 97US-0047639.  
PR 23-MAY-1997; 97US-0047640.  
PR 23-MAY-1997; 97US-0047641.  
PR 23-MAY-1997; 97US-0047642.  
PR 23-MAY-1997; 97US-0047643.  
PR 23-MAY-1997; 97US-0047644.  
PR 23-MAY-1997; 97US-0047645.  
PR 23-MAY-1997; 97US-0047646.  
PR 23-MAY-1997; 97US-0047647.  
PR 23-MAY-1997; 97US-0047648.  
PR 23-MAY-1997; 97US-0047649.  
PR 23-MAY-1997; 97US-0047650.  
PR 23-MAY-1997; 97US-0047651.  
PR 23-MAY-1997; 97US-0047652.  
PR 23-MAY-1997; 97US-0047653.  
PR 23-MAY-1997; 97US-0047654.  
PR 23-MAY-1997; 97US-0047655.  
PR 23-MAY-1997; 97US-0047656.  
PR 23-MAY-1997; 97US-0047657.  
PR 23-MAY-1997; 97US-0047658.  
PR 23-MAY-1997; 97US-0047659.  
PR 23-MAY-1997; 97US-0047660.  
PR 23-MAY-1997; 97US-0047661.  
PR 23-MAY-1997; 97US-0047662.  
PR 23-MAY-1997; 97US-0047663.  
PR 23-MAY-1997; 97US-0047664.  
PR 23-MAY-1997; 97US-0047665.  
PR 23-MAY-1997; 97US-0047666.  
PR 23-MAY-1997; 97US-0047667.  
PR 23-MAY-1997; 97US-0047668.  
PR 23-MAY-1997; 97US-0047669.  
PR 23-MAY-1997; 97US-0047670.  
PR 23-MAY-1997; 97US-0047671.  
PR 23-MAY-1997; 97US-0047672.  
PR 23-MAY-1997; 97US-0047673.  
PR 23-MAY-1997; 97US-0047674.  
PR 23-MAY-1997; 97US-0047675.  
PR 23-MAY-1997; 97US-0047676.  
PR 23-MAY-1997; 97US-0047677.  
PR 23-MAY-1997; 97US-0047678.  
PR 23-MAY-1997; 97US-0047679.  
PR 23-MAY-1997; 97US-0047680.  
PR 23-MAY-1997; 97US-0047681.  
PR 23-MAY-1997; 97US-0047682.  
PR 23-MAY-1997; 97US-0047683.  
PR 23-MAY-1997; 97US-0047684.  
PR 23-MAY-1997; 97US-0047685.  
PR 23-MAY-1997; 97US-0047686.  
PR 23-MAY-1997; 97US-0047687.  
PR 23-MAY-1997; 97US-0047688.  
PR 23-MAY-1997; 97US-0047689.  
PR 23-MAY-1997; 97US-0047690.  
PR 23-MAY-1997; 97US-0047691.  
PR 23-MAY-1997; 97US-0047692.  
PR 23-MAY-1997; 97US-0047693.  
PR 23-MAY-1997; 97US-0047694.  
PR 23-MAY-1997; 97US-0047695.  
PR 23-MAY-1997; 97US-0047696.  
PR 23-MAY-1997; 97US-0047697.  
PR 23-MAY-1997; 97US-0047698.  
PR 23-MAY-1997; 97US-0047699.  
PR 23-MAY-1997; 97US-0047700.  
PR 23-MAY-1997; 97US-0047701.  
PR 23-MAY-1997; 97US-0047702.  
PR 23-MAY-1997; 97US-0047703.  
PR 23-MAY-1997; 97US-0047704.  
PR 23-MAY-1997; 97US-0047705.  
PR 23-MAY-1997; 97US-0047706.  
PR 23-MAY-1997; 97US-0047707.  
PR 23-MAY-1997; 97US-0047708.  
PR 23-MAY-1997; 97US-0047709.  
PR 23-MAY-1997; 97US-0047710.  
PR 23-MAY-1997; 97US-0047711.  
PR 23-MAY-1997; 97US-0047712.  
PR 23-MAY-1997; 97US-0047713.  
PR 23-MAY-1997; 97US-0047714.  
PR 23-MAY-1997; 97US-0047715.  
PR 23-MAY-1997; 97US-0047716.  
PR 23-MAY-1997; 97US-0047717.  
PR 23-MAY-1997; 97US-0047718.  
PR 23-MAY-1997; 97US-0047719.  
PR 23-MAY-1997; 97US-0047720.  
PR 23-MAY-1997; 97US-0047721.  
PR 23-MAY-1997; 97US-0047722.  
PR 23-MAY-1997; 97US-0047723.  
PR 23-MAY-1997; 97US-0047724.  
PR 23-MAY-1997; 97US-0047725.  
PR 23-MAY-1997; 97US-0047726.  
PR 23-MAY-1997; 97US-0047727.  
PR 23-MAY-1997; 97US-0047728.  
PR 23-MAY-1997; 97US-0047729.  
PR 23-MAY-1997; 97US-0047730.  
PR 23-MAY-1997; 97US-0047731.  
PR 23-MAY-1997; 97US-0047732.  
PR 23-MAY-1997; 97US-0047733.  
PR 23-MAY-1997; 97US-0047734.  
PR 23-MAY-1997; 97US-0047735.  
PR 23-MAY-1997; 97US-0047736.  
PR 23-MAY-1997; 97US-0047737.  
PR 23-MAY-1997; 97US-0047738.  
PR 23-MAY-1997; 97US-0047739.  
PR 23-MAY-1997; 97US-0047740.  
PR 23-MAY-1997; 97US-0047741.  
PR 23-MAY-1997; 97US-0047742.  
PR 23-MAY-1997; 97US-0047743.  
PR 23-MAY-1997; 97US-0047744.  
PR 23-MAY-1997; 97US-0047745.  
PR 23-MAY-1997; 97US-0047746.  
PR 23-MAY-1997; 97US-0047747.  
PR 23-MAY-1997; 97US-0047748.  
PR 23-MAY-1997; 97US-0047749.  
PR 23-MAY-1997; 97US-0047750.  
PR 23-MAY-1997; 97US-0047751.  
PR 23-MAY-1997; 97US-0047752.  
PR 23-MAY-1997; 97US-0047753.  
PR 23-MAY-1997; 97US-0047754.  
PR 23-MAY-1997; 97US-0047755.  
PR 23-MAY-1997; 97US-0047756.  
PR 23-MAY-1997; 97US-0047757.  
PR 23-MAY-1997; 97US-0047758.  
PR 23-MAY-1997; 97US-0047759.  
PR 23-MAY-1997; 97US-0047760.  
PR 23-MAY-1997; 97US-0047761.  
PR 23-MAY-1997; 97US-0047762.  
PR 23-MAY-1997; 97US-0047763.  
PR 23-MAY-1997; 97US-0047764.  
PR 23-MAY-1997; 97US-0047765.  
PR 23-MAY-1997; 97US-0047766.  
PR 23-MAY-1997; 97US-0047767.  
PR 23-MAY-1997; 97US-0047768.  
PR 23-MAY-1997; 97US-0047769.  
PR 23-MAY-1997; 97US-0047770.  
PR 23-MAY-1997; 97US-0047771.  
PR 23-MAY-1997; 97US-0047772.  
PR 23-MAY-1997; 97US-0047773.  
PR 23-MAY-1997; 97US-0047774.  
PR 23-MAY-1997; 97US-0047775.  
PR 23-MAY-1997; 97US-0047776.  
PR 23-MAY-1997; 97US-0047777.  
PR 23-MAY-1997; 97US-0047778.  
PR 23-MAY-1997; 97US-0047779.  
PR 23-MAY-1997; 97US-0047780.  
PR 23-MAY-1997; 97US-0047781.  
PR 23-MAY-1997; 97US-0047782.  
PR 23-MAY-1997; 97US-0047783.  
PR 23-MAY-1997; 97US-0047784.  
PR 23-MAY-1997; 97US-0047785.  
PR 23-MAY-1997; 97US-0047786.  
PR 23-MAY-1997; 97US-0047787.  
PR 23-MAY-1997; 97US-0047788.  
PR 23-MAY-1997; 97US-0047789.  
PR 23-MAY-1997; 97US-0047790.  
PR 23-MAY-1997; 97US-0047791.  
PR 23-MAY-1997; 97US-0047792.  
PR 23-MAY-1997; 97US-0047793.  
PR 23-MAY-1997; 97US-0047794.  
PR 23-MAY-1997; 97US-0047795.  
PR 23-MAY-1997; 97US-0047796.  
PR 23-MAY-1997; 97US-0047797.  
PR 23-MAY-1997; 97US-0047798.  
PR 23-MAY-1997; 97US-0047799.  
PR 23-MAY-1997; 97US-0047800.  
PR 23-MAY-1997; 97US-0047801.  
PR 23-MAY-1997; 97US-0047802.  
PR 23-MAY-1997; 97US-0047803.  
PR 23-MAY-1997; 97US-0047804.  
PR 23-MAY-1997; 97US-0047805.  
PR 23-MAY-1997; 97US-0047806.  
PR 23-MAY-1997; 97US-0047807.  
PR 23-MAY-1997; 97US-0047808.  
PR 23-MAY-1997; 97US-0047809.  
PR 23-MAY-1997; 97US-0047810.  
PR 23-MAY-1997; 97US-0047811.  
PR 23-MAY-1997; 97US-0047812.  
PR 23-MAY-1997; 97US-0047813.  
PR 23-MAY-1997; 97US-0047814.  
PR 23-MAY-1997; 97US-0047815.  
PR 23-MAY-1997; 97US-0047816.  
PR 23-MAY-1997; 97US-0047817.  
PR 23-MAY-1997; 97US-0047818.  
PR 23-MAY-1997; 97US-0047819.  
PR 23-MAY-1997; 97US-0047820.  
PR 23-MAY-1997; 97US-0047821.  
PR 23-MAY-1997; 97US-0047822.  
PR 23-MAY-1997; 97US-0047823.  
PR 23-MAY-1997; 97US-0047824.  
PR 23-MAY-1997; 97US-0047825.  
PR 23-MAY-1997; 97US-0047826.  
PR 23-MAY-1997; 97US-0047827.  
PR 23-MAY-1997; 97US-0047828.  
PR 23-MAY-1997; 97US-0047829.  
PR 23-MAY-1997; 97US-0047830.  
PR 23-MAY-1997; 97US-0047831.  
PR 23-MAY-1997; 97US-0047832.  
PR 23-MAY-1997; 97US-0047833.  
PR 23-MAY-1997; 97US-0047834.  
PR 23-MAY-1997; 97US-0047835.  
PR 23-MAY-1997; 97US-0047836.  
PR 23-MAY-1997; 97US-0047837.  
PR 23-MAY-1997; 97US-0047838.  
PR 23-MAY-1997; 97US-0047839.  
PR 23-MAY-1997; 97US-0047840.  
PR 23-MAY-1997; 97US-0047841.  
PR 23-MAY-1997; 97US-0047842.  
PR 23-MAY-1997; 97US-0047843.  
PR 23-MAY-1997; 97US-0047844.  
PR 23-MAY-1997; 97US-0047845.  
PR 23-MAY-1997; 97US-0047846.  
PR 23-MAY-1997; 97US-0047847.  
PR 23-MAY-1997; 97US-0047848.  
PR 23-MAY-1997; 97US-0047849.  
PR 23-MAY-1997; 97US-0047850.  
PR 23-MAY-1997; 97US-0047851.  
PR 23-MAY-1997; 97US-0047852.  
PR 23-MAY-1997; 97US-0047853.  
PR 23-MAY-1997; 97US-0047854.  
PR 23-MAY-1997; 97US-0047855.  
PR 23-MAY-1997; 97US-0047856.  
PR 23-MAY-1997; 97US-0047857.  
PR 23-MAY-1997; 97US-0047858.  
PR 23-MAY-1997; 97US-0047859.  
PR 23-MAY-1997; 97US-0047860.  
PR 23-MAY-1997; 97US-0047861.  
PR 23-MAY-1997; 97US-0047862.  
PR 23-MAY-1997; 97US-0047863.  
PR 23-MAY-1997; 97US-0047864.  
PR 23-MAY-1997; 97US-0047865.  
PR 23-MAY-1997; 97US-0047866.  
PR 23-MAY-1997; 97US-0047867.  
PR 23-MAY-1997; 97US-0047868.  
PR 23-MAY-1997; 97US-0047869.  
PR 23-MAY-1997; 97US-0047870.  
PR 23-MAY-1997; 97US-0047871.  
PR 23-MAY-1997; 97US-0047872.  
PR 23-MAY-1997; 97US-0047873.  
PR 23-MAY-1997; 97US-0047874.  
PR 23-MAY-1997; 97US-0047875.  
PR 23-MAY-1997; 97US-0047876.  
PR 23-MAY-1997; 97US-0047877.  
PR 23-MAY-1997; 97US-0047878.  
PR 23-MAY-1997; 97US-0047879.  
PR 23-MAY-1997; 97US-0047880.  
PR 23-MAY-1997; 97US-0047881.  
PR 23-MAY-1997; 97US-0047882.  
PR 23-MAY-1997; 97US-0047883.  
PR 23-MAY-1997; 97US-0047884.  
PR 23-MAY-1997; 97US-0047885.  
PR 23-MAY-1997; 97US-0047886.  
PR 23-MAY-1997; 97US-0047887.  
PR 23-MAY-1997; 97US-0047888.  
PR 23-MAY-1997; 97US-0047889.  
PR 23-MAY-1997; 97US-0047890.  
PR 23-MAY-1997; 97US-0047891.  
PR 23-MAY-1997; 97US-0047892.  
PR 23-MAY-1997; 97US-0047893.  
PR 23-MAY-1997; 97US-0047894.  
PR 23-MAY-1997; 97US-0047895.  
PR 23-MAY-1997; 97US-0047896.  
PR 23-MAY-1997; 97US-0047897.  
PR 23-MAY-1997; 97US-0047898.  
PR 23-MAY-1997; 97US-0047899.  
PR 23-MAY-1997; 97US-0047900.  
PR 23-MAY-1997; 97US-0047901.  
PR 23-MAY-1997; 97US-0047902.  
PR 23-MAY-1997; 97US-0047903.  
PR 23-MAY-1997; 97US-0047904.  
PR 23-MAY-1997; 97US-0047905.  
PR 23-MAY-1997; 97US-0047906.  
PR 23-MAY-1997; 97US-0047907.  
PR 23-MAY-1997; 97US-0047908.  
PR 23-MAY-1997; 97US-0047909.  
PR 23-MAY-1997; 97US-0047910.  
PR 23-MAY-1997; 97US-0047911.  
PR 23-MAY-1997; 97US-0047912.  
PR 23-MAY-1997; 97US-0047913.  
PR 23-MAY-1997; 97US-0047914.  
PR 23-MAY-1997; 97US-0047915.  
PR 23-MAY-1997; 97US-0047916.  
PR 23-MAY-1997; 97US-0047917.  
PR 23-MAY-1997; 97US-0047918.  
PR 23-MAY-1997; 97US-0047919.  
PR 23-MAY-1997; 97US-0047920.  
PR 23-MAY-1997; 97US-0047921.  
PR 23-MAY-1997; 97US-0047922.  
PR 23-MAY-1997; 97US-0047923.  
PR 23-MAY-1997; 97US-0047924.  
PR 23-MAY-1997; 97US-0047925.  
PR 23-MAY-1997; 97US-0047926.  
PR 23-MAY-1997; 97US-0047927.  
PR 23-MAY-1997; 97US-0047928.  
PR 23-MAY-1997; 97US-0047929.  
PR 23-MAY-1997; 97US-0047930.  
PR 23-MAY-1997; 97US-0047931.  
PR 23-MAY-1997; 97US-0047932.  
PR 23-MAY-1997; 97US-0047933.  
PR 23-MAY-1997; 97US-0047934.  
PR 23-MAY-1997; 97US-0047935.  
PR 23-MAY-1997; 97US-0047936.  
PR 23-MAY-1997; 97US-0047937.  
PR 23-MAY-1997; 97US-0047938.  
PR 23-MAY-1997; 97US-0047939.  
PR 23-MAY-1997; 97US-0047940.  
PR 23-MAY-1997; 97US-0047941.  
PR 23-MAY-1997; 97US-0047942.  
PR 23-MAY-1997; 97US-0047943.  
PR 23-MAY-1997; 97US-0047944.  
PR 23-MAY-1997; 97US-0047945.  
PR 23-MAY-1997; 97US-0047946.  
PR 23-MAY-1997; 97US-0047947.  
PR 23-MAY-1997; 97US-0047948.  
PR 23-MAY-1997; 97US-0047949.  
PR 23-MAY-1997; 97US-0047950.  
PR 23-MAY-1997; 97US-0047951.  
PR 23-MAY-1997; 97US-0047952.  
PR 23-MAY-

PR 22-AUG-1997; 97US-0056880.  
 PR 22-AUG-1997; 97US-0056881.  
 PR 22-AUG-1997; 97US-0056882.  
 PR 22-AUG-1997; 97US-0056884.  
 PR 22-AUG-1997; 97US-0056886.  
 PR 22-AUG-1997; 97US-0056887.  
 PR 22-AUG-1997; 97US-0056888.  
 PR 22-AUG-1997; 97US-0056889.  
 PR 22-AUG-1997; 97US-0056892.  
 PR 22-AUG-1997; 97US-0056893.  
 PR 22-AUG-1997; 97US-0056900.  
 PR 22-AUG-1997; 97US-0056908.  
 PR 22-AUG-1997; 97US-0056909.  
 PR 22-AUG-1997; 97US-0056910.  
 PR 22-AUG-1997; 97US-0056911.  
 PR 05-SEP-1997; 97US-0057650.  
 PR 05-SEP-1997; 97US-0057669.  
 PR 05-SEP-1997; 97US-0057761.  
 PR 12-SEP-1997; 97US-0058785.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Bednarik DP, Brewer LA, Carter KC, Duan R, Edner R, Endress GA,  
 PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,  
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni Y, Olsen HS, Rosen CA,  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 WI: 1998-506364/43.  
 P-PSDB; AAW74777.

PT New isolated human genes and the secreted polypeptide(s) they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders

XX Claim 1; Page 279-280; 721pp; English.

XX This sequence represents a nucleic acid molecule designated Gene 47 from  
 CC the human cDNA clone HOGAV75 (deposited as clone ATCC 97899 and ATCC  
 CC 209045) which encodes a secreted human protein. The gene can be used to  
 CC generate fusion proteins by linking to the gene to a human  
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of  
 CC the fused protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new  
 CC polynucleotides. Specific uses are described for each of the 186  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAV59511 for described uses).

XX Sequence 1220 BP; 382 A; 207 C; 258 G; 372 T; 1 other;

XX Alignment\_scores:

Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

XX alignment\_block:

US-09-484-964-2 x AAV59557

Align seg 1/1 to: AAV59557 from: 1 to: 1220

19 GlyIunYrIlleYsLeuYsValIlleGlyGlnAspSerSerGluIleH 35  
 ||||||||||||||||||||||||||||||||||||||||||||  
 182 GGTGAATATATTAACTCAAGTCATGACAGATAGCAAGAGATTC 231  
 ||||||||||||||||||||||||||||||||||||||||||||  
 35 sPhelYsValIleYsLeuYsValIlleGlyGlnAspSerSerGluIleH 52  
 ||||||||||||||||||||||||||||||||||||||||||||  
 232 CTCGAAAGTGAATATGACACACATCTCAAGAAATCAAGATATACT 281

52 ysglnArggIngIyValPrometAsnSerLeuArgPheLeuPheGluGly 68  
 ||||||||||||||||||||||||||||||||||||||||||||  
 282 GTCGAAGACAGGGGTGTTCCATGAATTCACCTCAGTTCTCTTTCAGGCT 331  
 ||||||||||||||||||||||||||||||||||||||||||||  
 69 GluArgIleAlaAspAsnHisThrProlysgIuLeuGlyMetGluGlu 85  
 ||||||||||||||||||||||||||||||||||||||||||||  
 332 CACAGATTGCTGATATATCATCTCCAAAGAACTGGGAATGAGAGAGA 381  
 ||||||||||||||||||||||||||||||||||||||||||||  
 85 uaspValIlleGluValTYrGluGluGlnThrGlyGlyHisSerThrVal 101  
 ||||||||||||||||||||||||||||||||||||||||||||  
 382 AGATGATGATTGAAGTTATCAGAGAACAAAGGGGGCTATTCACACGTT 430  
 ||||||||||||||||||||||||||||||||||||||||||||

seq\_name: /SIDS1/gcdata/geneseq/geneseq/NA2000.DAT:AA235859

seq\_documentation\_block:

ID AA235859 standard; cDNA; 1514 BP.  
 AC AA235859;  
 XX

DT 03-FEB-2000 (first entry)

XX Human sentrin nucleotide sequence.

XX Human; sentrin; antisense oligonucleotide; phosphorothioate;  
 XX inhibition; modulation; expression; diagnosis; ss.

OS Homo sapiens.

XX Key

XX CDS

XX Location/Qualifiers

XX 136..441

XX /tag= a

XX /product= "sentrin"

XX US5985664-A.

XX 16-NOV-1999.

XX 17-DEC-1998; 98US-0213768.

XX 17-DEC-1998; 98US-0213768.

XX (ISIS-) ISIS PHARM INC.

XX Baker BF, Cowser LM;

XX WPI: 2000-022284/02.

XX P-PSDB; AAV49967.

XX Antisense compound which modulates human sentrin expression, useful for

XX treating diseases associated with sentrin expression -

XX Example 13: Column 39-42; 29pp; English.

XX The present invention describes an antisense compound (I) 8-30

XX nucleotides long targeted to a nucleic acid molecule encoding human

XX sentrin. The antisense compound comprises a phosphorothioate antisense

XX oligonucleotide which inhibits expression of human sentrin. (I) is

XX CC useful for inhibiting expression of sentrin in human cells or tissues

XX CC in vitro, for treating humans or other animals suspected of having or

XX CC being prone to a disease associated with sentrin expression. (I) can

XX CC also be used for research or diagnostic purposes. The present

XX Sequence 1514 BP; 462 A; 256 C; 314 G; 482 T; 0 other;

XX Alignment\_scores:

Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

XX alignment\_block:



US-09-484-964-2 x AAZ35859

Align seg 1/1 to: AAZ35859 from: 1 to: 1514

```

19 GlyGluTyrIleLeuLysLeuValIleGlyGlnAspSerSerGluIleH1 35
190 GGTGAATATATTAACCTCAAGATGATGACAGATAGCAGATGATCA 239
35 spheLysValLysMetThrThrHisLeuLysLysLeuGluSerTyrC 52
240 CTTCAAAGTGAATAATGACACATCTCAAGAACTCAAAAGATCATACT 289
52 ysgLnaArgGlnGlyValPrometAsnSerLeuArpPheLeuPheGluGly 68
290 GTCAAAAGACAGGGGTCTTCCATGATCACTACAGTTCTCTTGAGGCT 339
69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
340 CAGAGAATTCGTGATTAATCATCTCCAAAAGAACTGGGAATGGAGAGA 389
85 uaspValIleGluValTyrGlnGluGlnGlyHisSerThrVal 101
390 AGATGTGATTTGAATGATTTATCAGACAAACGCGGCTCATTCACACATT 438
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA16215
seq_documentation_block:
ID AAA16215 standard; DNA; 616 BP.
XX
XX AAA16215;
XX
XX 14-JUN-2000 (first entry)
XX
DE Human colon cancer differentially expressed nucleotide sequence #220.
XX
XX Colon cancer; detect; differential expression; human; treatment;
XX detect mutation; non-invasive diagnostic method; ds.
XX
XX Homo sapiens.
XX
XX MO200012702-A2.
XX
XX 09-MAR-2000.
XX
XX 30-AUG-1999; 99WO-US19424.
XX
XX 31-AUG-1998; 98US-0098639.
XX 27-JAN-1999; 99US-0117393.
XX
XX (FARB ) BAYER CORP.
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
XX Carino TJ, Dwiwedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
XX Schlegel R;
XX
XX WPI; 2000-256641/22.
XX
XX Novel nucleic acids and proteins for identifying therapeutic agents
XX useful for treating and diagnosing cancer, especially colon cancer
XX
XX Claim 16; Page 211-212; 345pp; English.
XX
XX This sequence represents a human nucleotide sequence which is
XX differentially expressed in colon cancer cells compared to the expression
XX levels in normal cells. The nucleotide sequence can be used as a source
XX of primers and probes. The nucleotide sequence is useful for determining
XX the phenotype of a cell by detecting the differential expression of the
XX sequence relative to a normal cell. The probes derived from the sequence
XX can also be used to determine the phenotype of cells in a sample. Probes
XX and antibodies which hybridize to the nucleotide sequence can also be
XX used to determine the phenotype of a cell. The primers are useful for
XX detecting a mutation in a test nucleotide sequence and also for detecting
XX cancer, preferably colon cancer. Antibodies against the protein encoded

```

CC by the nucleotide sequence can also be used in a method to detect colon  
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing  
 CC colon cancer at an early stage.

XX  
 XX Sequence 616 BP; 181 A; 109 C; 149 G; 170 T; 7 other;

## alignment\_scores:

Quality:	Length:
Ratio: 1.000	Gaps: 0
Percent Similarity: 100.000	Percent Identity: 100.000

## alignment\_block:

US-09-484-964-2 x AAA16215

Align seg 1/1 to: AAA16215 from: 1 to: 616

```

19 GlyGluTyrIleLeuLysLeuValIleGlyGlnAspSerSerGluIleH1 35
190 GGTGAATATATTAACCTCAAGATGATGACAGATAGCAGATGATCA 230
181 GGTGAATATATTAACCTCAAGATGATGACAGATAGCAGATGATCA 280
35 spheLysValLysMetThrThrHisLeuLysLysLeuGluSerTyrC 52
231 CTTCAAAGTGAATAATGACACATCTCAAGAACTCAAAAGATCATACT 280
52 ysgLnaArgGlnGlyValPrometAsnSerLeuArpPheLeuPheGluGly 68
281 GTCAAAAGACAGGGGTCTTCCATGATCACTACAGTTCTCTTGAGGCT 330
69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
331 CAGAGAATTCGTGATTAATCATCTCCAAAAGAACTGGGAATGGAGAGA 380
85 uaspValIleGlu 89
381 AGATGTGATTTGAA 393
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:AA63339
seq_documentation_block:
ID AAT63339 standard; CDNA; 372 BP.
XX
XX AAT63339;
XX
XX 17-AUG-1997 (first entry)
XX
XX Human host cell protein NPI-6 partial cDNA clone.
XX
XX Host cell protein; NPI-6; nucleoprotein interactor 6;
XX Influenza virus; replication; antiviral; virucide; ss.
XX
XX Homo sapiens.
XX
XX MO9712967-A1.
XX
XX 10-APR-1997.
XX
XX 06-OCT-1995; 95WO-US13044.
XX 06-OCT-1995; 95WO-US13044.
XX
XX (MOUN ) MOUNT SINAI MEDICAL CENT.
XX
XX Onelli R, Palese P;
XX
XX WPI; 1997-226211/20.
XX
XX New isolated DNA which encodes viral interacting proteins - used in
XX PT assays to isolate products for inhibiting viral protein binding
XX PT which is required for infection, replication, assembly or release
XX
XX Disclosure; Fig 11; 98pp; English.
XX

```

PI O'Neill R, Harty R, Palese PM;  
yy

PI O'Neill R, Harty R, Palese PM;  
xx



```

40 MetThrThrHisLeuLysLysLeuLysGluSerTyrCysGlnArgGlnI 56
   |||||||
153 ATGACACACGCAATCTCAAAAAGCTGAAAGAGTCATCTGTCAGAGACAGG 202
   |||||||
56 yValProMetAsnSerLeuArgPheLeuPheGluGluGlnArgIle 71
   |||||||
203 CSTTCCAATGAATTCTCTCAGGTTTGTGTTGAAGGCAAGAATC 248
   |||||||

```

OM of: US-09-484-964-2 to: Issued\_Patents\_NA:\* out\_format : pfs

Date: Aug 15, 2001 6:40 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

-MODEL=frame+p2n\_model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09484964/runat\_14082001\_111933\_27708/app\_query.fasta\_1.158  
-DB=Issued\_Patents\_NA -QPM=fastap -SUFFIX=oli.rni -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-DEIOP=6.000 -EGAEXT=7.000 -YGAPOP=60.000 -YGAEXT=60.000  
-DEIOP=6.000 -DEIEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=quality  
-THR\_MIN=20 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09484964 @CGNL\_1\_56 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPHY -WAIT -THREADS=1

#### Search information block:

Query: US-09-484-964-2  
Query Length: 101  
Database: Issued\_Patents\_NA:\*  
Database sequences: 324599  
Database length: 94655562  
Search time (sec): 79.730000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

#### score\_list:

Sequence /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:US-09-213-768-1 + 83.00 1549.08 1.1e-78 1514 1  
seq\_name: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:US-09-213-768-1

#### seq\_documentation\_block:

; Sequence 1, Application US/09213768  
; Patent No. 5985664  
; GENERAL INFORMATION:  
; APPLICANT: Iex M. Cowsett  
; APPLICANT: Brenda F. Baker  
; TITLE OF INVENTION: ANTISENSE MODULATION OF SENTRIN EXPRESSION  
; FILE REFERENCE: RTS-0026  
; CURRENT APPLICATION NUMBER: US/09/213,768  
; CURRENT FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 47  
; SEQ ID NO 1  
; LENGTH: 1514  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (136)..(441)  
; US-09-213-768-1

#### alignment\_scores:

Quality: 83.00 Length: 83  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-484-964-2 x US-09-213-768-1 ..

Align seg 1/1 to: US-09-213-768-1 from: 1 to: 1514

19 G1yG1uYr11e1yS1eulYsVal11eG1y1nAspSerSerg1u11eH1 35  
|||||  
190 GGTGAATATATTAACTCAAGTCATTGGACAGATAGCACTGAGATTCA 239

35 sPhelYsVal11yMetThrThrHis1eulYs1yS1eulYsG1uSerTYrC 52  
|||||  
240 CTCGAAGTGAATATGACACACATCTCAAGAACTCAAGATCATCT 289  
|||||  
52 ySg1nArG1nG1yVal1PrometaSnSer1eulArGPhelEuPheG1uG1y 68  
|||||  
290 GTCGAAGACAGGGGTTCATGAATTCATCAGGTTCTCTTGAGGGT 339  
|||||  
69 G1nArG11e1a1aSpasnH1sThrPro1ySg1u1eulYMetG1uG1uG1 85  
|||||  
340 CAGAGAATTGCTGATATATCTCTCCAAAAGAACTGGGATGAGGAGA 389  
|||||  
85 uAsPVal11eG1uValTYrG1nG1uG1nThrG1yG1yHisSerThrVal 101  
|||||  
390 AGATGTGATTGAAGTTTATCAGGAACAAGGGGGGTCTATTCAACAGTT 438  
|||||

Thu Aug 16 08:48:52 2001

us-09-484-964-2.oli.rni

---

Page 2

OM of: US-09-484-964-2 to: EST:\* out\_format : pfs

Date: Aug 15, 2001 6:21 PM

About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

MODEL=frame+g2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09484964/runal.14082001\_111932\_27678/app\_query.fasta\_1.158  
-DB=EST -QFMT=fastap -SUFFIX=ol1.rst -GAPOP=4.500 -GAPEXT=0.050  
-MINMATCH=0.050 -LOCPCL=0.000 -LOEPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.000 -XGAPOP=60.000 -XGAPEXT=60.000 -RGAPOP=6.000  
-RGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELCP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=ol1go -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=quality -THR.MIN=20 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pfs -NEM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US09484964\_ECGN1\_1\_2867 -NCPU=3 -LONGLOG -NO\_XLPHY  
-WAIT -THREADS=1

Search information block:

Query: US-09-484-964-2

Query length: 101

Database: EST\*

Database sequences: 10228115

Database length: 431459454

Search time (sec): 1164.220000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
gb_est8:AA473590	+	83.00	1689.78	5.3e-85	321	AA473590 v63h04.r1 Beddington
gb_est4:AA227354	+	83.00	1688.94	5.9e-85	361	AA227354 zrz2ell.r1 Stratagene
gb_est4:AA227355	+	83.00	1688.80	6.0e-85	366	AA227355 zrz2ell.r1 Stratagene
gb_est81:BE982210	-	83.00	1688.70	6.1e-85	373	BE982210 UT-M-CG0P-bf1-d-02-0-U
gb_est17:AT180895	+	83.00	1688.42	6.3e-85	388	AT180895 ub77ad5.r1 Soares_mamm
gb_est14:AA981271	+	83.00	1688.31	6.3e-85	391	AA981271 vx60f17.r1 Stratagene
gb_est1:AA421208	+	83.00	1688.16	6.5e-85	402	AA421208 u050h05.r1 Soares_test
gb_est11:AA726795	+	83.00	1688.00	6.6e-85	407	AA726795 v042910.r1 Barstead mc
gb_est18:AA507248	-	83.00	1687.66	6.9e-85	431	AA507248 v178009.r1 Barstead mc
gb_est88:BF512106	-	83.00	1687.85	6.8e-85	411	BF512106 UT-H-BMI-am1-g-09-0-UT
gb_est17:AT180896	-	83.00	1687.85	6.8e-85	420	AT180896 ub77ad5.r1 Soares_mamm
gb_est13:AA896203	+	83.00	1687.74	6.9e-85	426	AA896203 vx62d07.r1 Stratagene
gb_est15:AT047190	+	83.00	1687.69	6.9e-85	429	AT047190 u062d04.r1 Soares_embd
gb_est15:BF978928	+	83.00	1687.68	6.9e-85	430	BF978928 602147602p1 NIH_MGC_62
gb_est10:AA682161	+	83.00	1687.66	6.9e-85	431	AA682161 v013e09.r1 Barstead mc
gb_est110:W24066	+	83.00	1687.63	7.0e-85	433	W24066 z082d02.r1 Soares_sensc
gb_est7:AA423388	+	83.00	1687.61	7.0e-85	434	AA423388 vx39b12.r1 Soares_mamm
gb_est13:AT1646104	-	83.00	1687.53	7.1e-85	438	AT1646104 ub77ad5.r1 Soares_mamm
gb_est16:BF333866	+	83.00	1687.53	7.1e-85	439	BF333866 SNEST420F08.Y1 cs1
gb_est10:AA636685	+	83.00	1687.48	7.1e-85	442	AA636685 v15f02.r1 Barstead mc
gb_est15:BE482568	+	83.00	1687.44	7.1e-85	444	BE482568 168404 BARC_5BOV_Bos t
gb_est1:AA401864	+	83.00	1687.36	7.2e-85	449	AA401864 zve5b12.r1 Soares_totd
gb_est17:BE629376	+	83.00	1687.35	7.2e-85	450	BE629376 u037c02.r1 Soares_mamm
gb_gssl:AA046515	+	83.00	1687.35	7.2e-85	450	AA046515 AF046515 Mus musculus
gb_est106:NJ31695	+	83.00	1687.25	7.3e-85	456	NJ31695 yx69g01.r1 Soares_melanc
gb_est1:AA105474	+	83.00	1687.21	7.4e-85	459	AA105474 mm32d10.r1 Stratagene
gb_est13:BE331886	+	83.00	1687.17	7.4e-85	461	BE331886 ut05c09.y1 Soares_thym
gb_est17:BE629396	+	83.00	1687.10	7.5e-85	466	BE629396 u037d12.y1 Soares_mamm
gb_est15:BE482995	+	83.00	1687.02	7.5e-85	471	BE482995 168990 BARC_5BOV_Bos t
gb_est17:BE197453	+	83.00	1686.87	7.7e-85	481	BE197453 u074f06.y1 Soares_mamm
gb_est148:AW519530	+	83.00	1686.75	7.8e-85	489	AW519530 x033h12.y1 Soares_mous
gb_est1:AA030059	+	83.00	1686.72	7.8e-85	491	AA030059 m122h03.r1 Soares_mous
gb_est17:AA399812	+	83.00	1686.72	7.8e-85	491	AA399812 v07f1f12.r1 Beddington
gb_est15:AW323392	+	83.00	1686.72	7.8e-85	491	AW323392 u038g05.y1 NCI_CGAP_Lu
gb_est17:AT1248769	-	83.00	1686.65	7.9e-85	496	AT1248769 qh77h07.x1 Soares_feta
gb_est14:AW239296	+	83.00	1686.65	7.9e-85	496	AW239296 x038e11.y1 NCI_CGAP_Lu
gb_est14:AW989449	+	83.00	1686.65	7.9e-85	496	AW989449 u0414c01.y1 Soares_mamm
gb_est1:AW740709	+	83.00	1686.57	8.0e-85	501	AW740709 u050b06.y1 NCI_CGAP_Me
gb_est12:BE226598	+	83.00	1686.57	8.0e-85	501	BE226598 1a23c07.y1 Mouse_E10_5

seq_name: gb_est8:AA473590	seq_documentation_block:	LOCUS	AA473590	321 bp	mRNA	EST	18-JUN-1997
gb_est7:AA399819	+	83.00	1686.32	8.2e-85	519	AA399819 v071h12.r1 Beddington	
gb_est80:BE861560	+	83.00	1686.11	8.5e-85	534	BE861560 UT-M-AH1-ago-e-04-0	
gb_est85:BF244251	+	83.00	1686.06	8.5e-85	538	BF244251 601863007F1 NIH_MGC	
gb_est40:AV667206	+	83.00	1686.86	8.7e-85	553	AV667206 AV667206 Bos tauus	
gb_est73:BE374647	+	83.00	1686.85	8.8e-85	554	BE374647 601224145F1 NCI_CGA	

seq\_name: gb\_est8:AA473590

seq\_documentation\_block:

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

FEATURES

source

1. 321

/organism="Mus musculus"

/strain="C57BL6 x DBA"

/db\_xref="taxon:10090"

/clone="IMAGE:866071"

/clone\_lib="Beddington mouse embryonic region"

/sex="pooled"

/tissue\_type="embryo"

/dev\_stage="7.5dpc"

/lab\_host="DH12S"

/note="Organ: whole embryo; Vector: pCMV-SPORT; Site: 1; SalI: Site 2; NotI: Cloned unidirectionally. Primer: Oligo dt. Gastrulating embryos were collected at 7.5dpc from C57BL6 x DBA matings, excluding embryos that had developed head folds and all extraembryonic tissues. Average insert size: 1.3 kb (range: 0.5 - 3.0 kb). Referenced in Development 121, 2479-2489 (1995)."

BASE COUNT

ORIGIN

alignment\_scores:

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-09-484-964-2 x AA473590

Align seg 1/1 to: AA473590 from: 1 to: 321

19 GlycyltyrilelyleuylsValIleGlylnAspserSerGluIleH 35  
 56 GGAGAAATCAATTAACCAAGTTATGGACAGATAGCGATGATAC 105  
 35 sPhelysVallysmethrThrHisLeuylsLeuylsGluSerTyrC 52  
 106 TTTCAGAGTGAATAATGACACATCTCCAGAACTCAAGAAATCACTACT 155  
 52 ysglnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheGluGly 68  
 156 GTCAAAAGACAGGAGTTCATGAATTCAGTTCTCTTCTTGAAGG 205  
 69 GluArgIleAlaAspAsnHisThrProlysgLuleuGlymetLuleuG 85  
 206 CAGAGAAATTCCTGATATCACTATCCGAAAGAACTGGAAATGAGAGA 255  
 85 uaspyAlIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThrVal 101  
 256 AGATGTGATTTGAAGTTTATCAGAACAAACGGGGGTCACTCAGCGTT 304

seq\_name: gb\_est4:AA227354

seq\_documentation\_block:

LOCUS AA227354 361 bp mRNA EST 24-FEB-1997  
 DEFINITION zr22e12.r1 StrataGene NT2 neuronal precursor 937230 Homo sapiens  
 CDNA clone IMAGE:664172 5' similar to TR:G927779 G927779 SUPPRESSOR  
 OF MIF2 MUTATIONS. ;, mRNA sequence.

ACCESSION AA227354  
 VERSION AA227354.1 GI:1848909  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 361)

REFERENCE

AUTHORS

Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
 Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins  
 ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Merdis,E., Moore  
 Schellenberg,K., Soares,M.B., Tan,F., Thierly-Meg,J., Treviskis,E.,  
 Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)

JOURNAL MEDLINE

COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 This clone is available royally-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 282.

FEATURES

source

1. 361  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5426089"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:664172"  
 /clone\_lib="Stratagene NT2 neuronal precursor 937230"  
 /tissue\_type="neuroepithelial cells"  
 /dev\_stage="Ntera-2 neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: brain; Vector: pBluescript SK-; Site:1:  
 EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. Uninduced, exponentially growing neuroepithelial  
 cells (Ntera-2/c1.D1). Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGACGACG  
 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' "  
 121 a 68 c 92 g 80 t

alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-484-964-2 x AA227354 ..

Align seg 1/1 to: AA227354 from: 1 to: 361

19 GlycyltyrilelyleuylsValIleGlylnAspserSerGluIleH 35  
 108 GGTGAATATTAATTAACCAAGTTATGGACAGATAGCGATGATAC 157  
 35 sPhelysVallysmethrThrHisLeuylsLeuylsGluSerTyrC 52  
 158 CTCAAGTGAATAATGACACATCTCCAGAACTCAAGAAATCACTACT 207  
 52 ysglnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheGluGly 68  
 208 GTCAAAAGACAGGAGTTCATGAATTCACACAGTTCTCTTGAAGGT 257  
 69 GluArgIleAlaAspAsnHisThrProlysgLuleuGlymetLuleuG 85  
 258 CAGAGAAATTCCTGATATCACTATCCAGAAAGAACTGGAAATGAGAGA 307  
 85 uaspyAlIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThrVal 101  
 308 AGATGTGATTTGAAGTTTATCAGAACAAACGGGGGTCACTCAGCGTT 356

seq\_name: gb\_est4:AA227355

seq\_documentation\_block:

LOCUS AA227355 368 bp mRNA EST 11-MAR-1998  
 DEFINITION zr22e12.r1 StrataGene NT2 neuronal precursor 937230 Homo sapiens  
 CDNA clone IMAGE:664174 5' similar to TR:G927779 G927779 SUPPRESSOR  
 OF MIF2 MUTATIONS. ;, mRNA sequence.

ACCESSION AA227355  
 VERSION AA227355.1 GI:1848910  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 368)

REFERENCE

AUTHORS

Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin  
 ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
 White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST Project  
 Unpublished (1997)

JOURNAL

COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 This clone is available royally-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1260 Std Error: 0.00  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 360.

FEATURES

source

1. 368  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5426092"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:664174"  
 /clone\_lib="Stratagene NT2 neuronal precursor 937230"  
 /tissue\_type="neuroepithelial cells"  
 /dev\_stage="Ntera-2 neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"





REFERENCE Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 388)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:906196

FEATURES Possible reversed clone; similarity on wrong strand  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 185.

source Location/Qualifiers  
1..388  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1383728"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DHI0B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 129 a 72 c 103 g 82 t 2 others

ORIGIN

alignment\_scores: Quality: 83.00 Length: 83  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block: US-09-484-964-2 x A1180895 ..

Align seg 1/1 to: A1180895 from: 1 to: 388

```

19 G1yG1uTyr1leUys1euYsVal1leG1yG1nAsPseSersG1u1leH1 35
|||||
140 GGAGAAATACATTAACCTAAAGTTATTGGACAGATAGCAGTACATACA 189
|||||
35 sPhelysVal1ysMeThrThrHis1eUys1s1eUys1G1uSertTyrC 52
|||||
190 TTTCAAGTAAATGACACACATCTCAAGAAATCAAGAAATCAATCTACT 239
|||||
52 ySg1nArG1ng1yVal1PrOmEtAsnSer1euArGpHeUeUpHeG1uG1y 68
|||||
240 GTCAAAGACAGGAGGATTCATGATTCACACTGCTTCTCTTTGAAGT 289
|||||
69 G1nArG1leAlaAsPasnH1stHrPrOlySg1uLeUg1yMeTc1uG1uG1 85
|||||
290 CAGAGATTCCTGATATCATCTCCGAAAGAACTGGGAATGAGAGAGA 339
|||||
85 uAsPVal1leG1uVal1yrrG1ng1uG1nThrG1yG1yH1SertThrVal 101
|||||
340 AGATGATGATTGAAGTTTATCAGGACAAACGGGGGCTCAGTCGAGGTT 388

```

seq\_name: gb\_est14:AA981271

seq\_documentation\_block: 391 bp mRNA EST 27-MAY-1998

LOCUS AA981271

DEFINITION vx60f07.r1 Strataegene mouse macrophage (#937306) Mus musculus cDNA

clone IMAGE:1279621 5' similar to SW:SM33\_HUMAN Q93068

UBIQUITIN-LIKE PROTEIN SMT3C ; mRNA sequence.

ACCESSION AA981271

VERSION AA981271.1 GI:3159807

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 391)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:671421

Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 379.

source Location/Qualifiers  
1..391  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1279621"  
/clone\_lib="Strataegene mouse macrophage (#937306)"  
/tissue\_type="macrophage"  
/dev\_stage="WEHI-3 cell line"  
/note="Organ: blood. Vector: pBluescript SK-; Site 1: EcoRI. Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. WEHI-3 cell line. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 128 a 70 c 95 g 97 t 1 others

ORIGIN

alignment\_scores: Quality: 83.00 Length: 83  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block: US-09-484-964-2 x AA981271 ..

Align seg 1/1 to: AA981271 from: 1 to: 391

```

19 G1yG1uTyr1leUys1euYsVal1leG1yG1nAsPseSersG1u1leH1 35
|||||
114 GGAGATACATTAACCTAAAGTTATTGGACAGATAGCAGTACATACA 163
|||||
35 sPhelysVal1ysMeThrThrHis1eUys1s1eUys1G1uSertTyrC 52
|||||
164 TTTCAAGTAAATGACACACATCTCAAGAAATCAAGAAATCAATCTACT 213
|||||
52 ySg1nArG1ng1yVal1PrOmEtAsnSer1euArGpHeUeUpHeG1uG1y 68
|||||
214 GTCAAAGACAGGAGGATTCATGATTCACAGTTCCTCTTTGAAGT 263

```

```

69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
|||||
264 CAGAGAAATTCCTGATATCATCTACCCGAAAGAACTGGCAATGAGAGAGA 313
85 uasppvalli1leGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
|||||
314 AGATGATTTGAAGTTATTCAGAACAAACGGGGGCTCACTGACAGGTT 362

seq_name: gb_est1:AA421208

seq_documentation_block:
LOCUS AA421208 402 bp mRNA EST 16-OCT-1997
DEFINITION Zu05h05.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:731001
5' similar to TR:G927779 G927779 SUPPRESSOR OF MIF2 MUTATIONS. ;
mRNA sequence.
ACCESSION AA421208
VERSION AA421208.1 GI:2100033
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 402)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
T., Waterston, R., Wilson, R.
WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson R
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 391.
Location/Qualifiers
1. 402
/organism="Homo sapiens"
/db_xref="GDB:5927801"
/db_xref="taxon:9606"
/clone_lib="IMAGE:731001"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGCGCCATCTTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 134 a 74 c 105 g 89 t
ORIGIN
alignment_scores:
Quality: 83.00 Length: 83
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-484-964-2 x AA421208 ..
Align seg 1/1 to: AA421208 from: 1 to: 402

```

```

19 GlnGluTyrIleLysLeuLysValIleGlyGlnAspSerGluIleH1 35
|||||
147 GGTGATATATTAACATAAGTCAATGGACAGGATAGCAGAGATTTCA 196
35 sPhelysValLysMetLthrThHisLysLysLysLysGluSerTyrC 52
|||||
197 CTTCAAGATGAAATGACANACACATCTCAAGAACTCAAGAAATCATACT 246
52 ysgInArgGlnGlnValPrometAsnSerLeuArgPheLeuPheLugly 68
|||||
247 GTCAAGACAGAGGTTCCATGAATCAATCACTCAGGTTCTCTTGAAGGT 296
69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
|||||
297 CAGAGAAATTCCTGATATCATCTACCCGAAAGAACTGGCAATGAGAGAGA 346
85 uasppvalli1leGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
|||||
347 AGATGATTTGAAGTTATTCAGAACAAACGGGGGCTCACTGACAGGTT 395

seq_name: gb_est1:AA726795

seq_documentation_block:
LOCUS AA726795 402 bp mRNA EST 02-JAN-1998
DEFINITION vu42q10.r1 Barstead mouse myotubes MRLB5 Mus musculus cDNA clone
IMAGE:1194114 5' similar to SW:SM33_HUMAN Q93068 UBQUITIN-LIKE
PROTEIN SMT3C ; mRNA sequence.
ACCESSION AA726795
VERSION AA726795.1 GI:2744502
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
REFERENCE 1 (bases 1 to 402)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:641210
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 385.
Location/Qualifiers
1. 402
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone_lib="IMAGE:1194114"
/clone_lib="Barstead mouse myotubes MRLB5"
/cell_line="C2C12"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGCGCCATCTTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(PAATCGCATCTCTG), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing

```

BASE COUNT 136 a 72 c 104 g 90 t  
 ORIGIN

alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-484-964-2 x AA726795 ..

Align seg 1/1 to: AA726795 from: 1 to: 402

19 GlyGluTyrIleLeuLysValIleGlyGlnAspSerSergIuIleH1 35  
 139 GGAGATACATTAACCTCAAGTTATTGGACAGATAGCAGTGAATAC 188  
 35 sPhelysValLysMetThrHisLeuLysLysLeuLysGluSerTyrC 52  
 189 TTTCAAGTGAATAATGACACATCTCAAGAACTCAAGATCATCTACT 238  
 52 YSGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuGluGly 68  
 239 GTCAAGACAGGAGTTCATGATTCCTCAGGTTCTCTTGAAGGT 288  
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85  
 289 CAGAGATTGCTGATATCATCTCCGAAGAAGCTGGAATGAGAGACA 338  
 85 uAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThrVal 101  
 339 AGATGGATTTGAAGTTATCAGGACAAACGGGGGCTCCTCAGCGGT 387

seq\_name: gb\_est7:AA450748

seq\_documentation\_block:

LOCUS AA450748 407 bp mRNA EST 04-JUN-1997  
 DEFINITION V178604.1 Soares\_mammary\_gland\_NBMWG Mus musculus cDNA clone  
 IMAGE:849895 5' similar to TR:G927779 G927779 SUPPRESSOR OF MIF2  
 MUTATIONS: ; mRNA sequence.

ACCESSION AA450748

VERSION AA450748.1 GI:2164418

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE  
 AUTHORS Mairra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucada,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The Masht-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Mairra M/Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

FEATURES  
 SOURCE

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LIND; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:502047

Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 387.

Location/Qualifiers  
 1. 407  
 /organism="Mus musculus"  
 /strain="C57BL/6J"

/db\_xref="taxon:10090"  
 /clone="IMAGE:849895"  
 /clone\_1ib="Soares\_mammary\_gland\_NBMWG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia  
 ) with a modified polylinker; Site 1: Not I; Site 2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5',  
 TGTTCACATCTGAAAGTGGAGCCGCCGAATGCTTTTCTTTTCTTTTCTTTT  
 T 3'] double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M. Fatima  
 Bonaldo."

BASE COUNT 134 a 72 c 104 g 97 t  
 ORIGIN

alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-484-964-2 x AA450748 ..

Align seg 1/1 to: AA450748 from: 1 to: 407

19 GlyGluTyrIleLeuLysValIleGlyGlnAspSerSergIuIleH1 35  
 137 GGAGATACATTAACCTCAAGTTATTGGACAGATAGCAGTGAATAC 186  
 35 sPhelysValLysMetThrHisLeuLysLysLeuLysGluSerTyrC 52  
 187 TTTCAAGTGAATAATGACACATCTCAAGAACTCAAGATCATCTACT 236  
 52 YSGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuGluGly 68  
 237 GTCAAGACAGGAGTTCATGATTCCTCAGGTTCTCTTGAAGGT 286  
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85  
 287 CAGAGATTGCTGATATCATCTCCGAAGAAGCTGGAATGAGAGACA 336  
 85 uAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThrVal 101  
 337 AGATGGATTTGAAGTTATCAGGACAAACGGGGGCTCCTCAGCGGT 385

seq\_name: gb\_est88:BF512106

seq\_documentation\_block:

LOCUS BF512106 411 bp mRNA EST 07-DEC-2000  
 DEFINITION UI-H-BW1-emi-9-09-0-UI.s1 NCI\_CGAP\_Sub7 Homo sapiens cDNA clone  
 IMAGE:3070264 3', mRNA sequence.

ACCESSION BF512106

VERSION BF512106.1 GI:11597318

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 The sequence contained an oligo-dT track that was present in the



US-09-484-964-2 x A1180896

Align seg 1/1 to: A1180896 from: 1 to: 420

19 GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH1 35  
 143 GGAGATACATTAACCTCAAGATTATGGACAGATAGCATGATACA 192  
 35 sPhelysValLysMetThrThrHisLeuLysLysLysGluSerTyrC 52  
 193 TTTCAAGTCAAAATGACACACATCTCAAGAAATCAAGATCATACT 242  
 52 ysglnarvgngllyvalprometasenleuargpheleupheglugly 68  
 243 GTCAAGAGACAGGAGTTCATTAATTCACAGGTTTCTTTGAAGGT 292  
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85  
 293 CAGAGAAATGCTGATATCATCTCCGAAAGAACTGGAAATGAGAGAGA 342  
 85 uaspvalilegluvaltyrvgngluginthrlyglhisserthrval 101  
 343 AGATGTGATTGAAGTTTATCAGAGAACAAAGGGGGGTCTCAGCGGT 391

seq\_name: gb\_est13:AA896203

seq\_documentation\_block:

LOCUS AA896203 426 bp mRNA EST 06-APR-1998  
 DEFINITION vx2d07.r1 StrataGene mouse macrophage (#937306) Mus musculus CDNA

clone IMAGE:1279789 5' similar to SW:SM33\_HUMAN Q93068  
 UBIQUITIN-LIKE PROTEIN SMT3C ; mRNA sequence.

ACCESSION AA896203  
 VERSION AA896203.1 GI:3032596

KEYWORDS EST.  
 SOURCE house mouse.

## ORGANISM

Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 426)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HHMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

## REFERENCE

## AUTHORS

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:671589  
 Seq primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 407.

## FEATURES

## SOURCE

1..426

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="IMAGE:1279789"

/clone\_lib="StrataGene mouse macrophage (#937306)"

/tissue\_type="macrophage"

/dev\_stage="WEHI-3 cell line"

/lab\_host="SOLR (kanamycin resistant)"

/note="Organ blood; Vector: pBluescript SK-; Site:1;

EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. WEHI-3 cell line. Average insert size: 1.5 kb;

Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGCGACGAG

3' ~3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT

139 a 80 c 114 g 93 t

## ORIGIN

alignment\_scores:  
 Quality: 83.00 Length: 83  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-484-964-2 x AA896203

Align seg 1/1 to: AA896203 from: 1 to: 426

19 GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH1 35  
 170 GGAGATACATTAACCTCAAGATTATGGACAGATAGCATGATACA 219  
 35 sPhelysValLysMetThrThrHisLeuLysLysLysGluSerTyrC 52  
 220 TTTCAAGTCAAAATGACACACATCTCAAGAAATCAAGATCATACT 269  
 52 ysglnarvgngllyvalprometasenleuargpheleupheglugly 68  
 270 GTCAAGAGACAGGAGTTCATTAATTCACAGGTTTCTTTGAAGGT 319  
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85  
 320 CAGAGAAATGCTGATATCATCTCCGAAAGAACTGGAAATGAGAGAGA 369  
 85 uaspvalilegluvaltyrvgngluginthrlyglhisserthrval 101  
 370 AGATGTGATTGAAGTTTATCAGAGAACAAAGGGGGGTCTCAGCGGT 418

seq\_name: gb\_est15:AI047190

seq\_documentation\_block:

LOCUS AI047190 429 bp mRNA EST 08-JUL-1998  
 DEFINITION uh62d04.r1 Soares embryonic stem cell NMES Mus musculus CDNA clone

IMAGE:1749991 5' similar to SW:SM33\_HUMAN Q93068 UBIQUITIN-LIKE  
 PROTEIN SMT3C ; mRNA sequence.

ACCESSION AI047190  
 VERSION AI047190.1 GI:3295477

## KEYWORDS

## SOURCE

## ORGANISM

house mouse.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 429)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

## REFERENCE

## AUTHORS

TITLE The WashU-HHMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

## REFERENCE

## AUTHORS

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:961803  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 407.

## FEATURES

## SOURCE

1..429

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="IMAGE:1749991"

/clone\_lib="Soares\_embryonic\_stem\_cell\_NMES"

```

/cell_type="embryonic stem cell"
/lab_host="DH10B"
/note="Organ: bone marrow; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'
TCTTGCACATCTGAAGTGGAGCGCCGATGTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through two rounds of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      140 a      76 c      109 g      104 t
ORIGIN
alignment_scores:
  quality:      83.00      length:      83
  ratio:        1.000      gaps:        0
  percent similarity: 100.000      percent identity: 100.000
alignment_block:
  US-09-484-964-2 x AI047190
Align seg 1/1 to: AI047190 from: 1 to: 429
19 GYGLUTYRILEYSLYSLYVAIIIEGLYGLNAPSERSERGIIIEH 35
|||||
154 GGAGAAATACATTAACTCAAGATTGACAGAGATGAGTGATGATCA 203
35 SPHELYSVALISMETHRTHRHISLEULYSLEULYSGLUSERTYRC 52
|||||
204 TTTCAAAGTGAAATGACACACATCTCAAGAAATCAAAATCATACT 253
52 YSGLNARGINGLYVALPROMETASNSERLEUARGPHELEUPHEGLUG 68
|||||
254 GTCAAAAGCAGGAGTTCACATGAAATTCACCTCAGGTTCTTTGAAGT 303
69 GINARGILEALASPASNHSIHPROLYSGIULEUGLYMETGLUGI 85
|||||
304 CAGGAATTCGTGATATCATCTCCGAAAGAACTGGGAATGGAGAGA 353
85 UASPYAIIIEGLUVALTYRGINGLUGINTHRIGLYGLYHISSETRHVAL 101
|||||
354 AGATGATTTGAAGTTTATCAGGAAACAGGGGGTCTCAGCGGTT 402
seq_name: gb_est195:BF978928
seq_documentation_block:
  LOCUS      BF978928      430 bp      mRNA      EST      22-JAN-2001
  DEFINITION  602147602F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4306489 5',
  ACCESSION  BF978928
  VERSION    BF978928.1 GI:12346143
  KEYWORDS   EST.
  SOURCE      human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  REFERENCE  1 (bases 1 to 430)
              NIH-MGC http://mgi.nci.nih.gov/.
  AUTHORS    National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE      Unpublished (1999)
  JOURNAL    Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: ATCC/DCTD/DTF
              CDNA Library Preparation: CLONTECH Laboratories, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L1CM1175 row: c column: 02

```

```

FEATURES             High quality sequence stop: 430.
  source              Location/Qualifiers
                      1..430
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4306489"
                      /clone_lib="NIH_MGC_62"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /lab_type="melanotic melanoma, high MDR"
                      /note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
                      SfiI (ggccgcctggcc); Site_2: SfiI (ggccatctggcc);
                      Double-stranded cDNA was prepared from cell line RNA. 5'
                      and 3' adaptors were used in cloning as follows: 5'
                      adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor
                      sequence: 5'-ATTCTAGAGCGCGGCGGAGCATG-dt(30)BN-3'
                      (where B = A, C, or G and N = A, C, G, or T). Average
                      insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                      contained inserts by PCR. This library was enriched for
                      full-length clones and was constructed by Clontech
                      Laboratories (Palo Alto, CA)."
BASE COUNT      134 a      80 c      119 g      97 t
ORIGIN
alignment_scores:
  quality:      83.00      length:      83
  ratio:        1.000      gaps:        0
  percent similarity: 100.000      percent identity: 100.000
alignment_block:
  US-09-484-964-2 x BF978928
Align seg 1/1 to: BF978928 from: 1 to: 430
19 GYGLUTYRILEYSLYSLYVAIIIEGLYGLNAPSERSERGIIIEH 35
|||||
171 GGATGATATATTAACTCAAGATTCAGACAGATGACAGATATCA 220
35 SPHELYSVALISMETHRTHRHISLEULYSLEULYSGLUSERTYRC 52
|||||
221 CTTCAAGAGTGAAGAAATGACAAACATCTCAAGAAACTCAAAATCATACT 270
52 YSGLNARGINGLYVALPROMETASNSERLEUARGPHELEUPHEGLUG 68
|||||
271 GTCAAAAGCAGGAGTTCACATGAAATTCACCTCAGGTTCTTTGAAGT 320
69 GINARGILEALASPASNHSIHPROLYSGIULEUGLYMETGLUGI 85
|||||
321 CAGGAATTCGTGATATCATCTCCAAAAGAACTGGGAATGGAGAGA 370
85 UASPYAIIIEGLUVALTYRGINGLUGINTHRIGLYGLYHISSETRHVAL 101
|||||
371 AGATGATTTGAAGTTTATCAGGAAACAGGGGGTCTCAGCGGTT 419
seq_name: gb_est10:AA682161
seq_documentation_block:
  LOCUS      AA682161      431 bp      mRNA      EST      05-DEC-1997
  DEFINITION  v013609.r1 Bartshead mouse myotubes MRLRB5 Mus musculus cDNA clone
  IMAGE:1180552 5', similar to TR:G927779 G927779 SUPPRESSOR OF MIF2
  ACCESSION  AA682161
  VERSION    AA682161.1 GI:2663193
  KEYWORDS   EST.
  SOURCE      house mouse.
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE  1 (bases 1 to 431)
              Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
              Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
              Schellenberg, K., Stepien, M., Tan, F., Underwood, R., Moore, B.,
              Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

```

